



Implementación de una estrategia *in-silico* para la identificación de péptidos candidatos a vacuna terapéutica individualizada en tumor de paciente con PEComa

Diego Alfredo Amaya Ramírez

Universidad Nacional de Colombia
Facultad de Medicina - Facultad de Ingeniería
Bogotá, Colombia
2019

Implementación de una estrategia *in-silico* para la identificación de péptidos candidatos a vacuna terapéutica individualizada en tumor de paciente con PEComa

Diego Alfredo Amaya Ramírez

Tesis de investigación presentada como requisito parcial para optar al título de:

Magister en Ingeniería Biomédica

Director:

Luis Fernando Niño Vásquez, Ph. D.

Codirector:

Carlos Alberto Parra López, Ph. D.

Línea de Investigación:

Diseño de vacunas terapéuticas para cáncer basadas en péptidos

Grupos de Investigación:

Laboratorio de Investigación en Sistemas Inteligentes (LISI) - Luis Fernando Niño Vásquez

Inmunología y Medicina Traslacional (I&MT) - Carlos Alberto Parra López

Universidad Nacional de Colombia

Facultad de Medicina - Facultad de Ingeniería

Bogotá D. C., Colombia

2019

Este logro se lo dedico a mi madre, por todos sus esfuerzos y sacrificios que ha hecho y sigue haciendo para sacar su Familia adelante.

« If I have seen further, it is by standing upon the shoulders of giants »

Isaac Newton

« Concern for man and his fate must always form the chief interest of all technical endeavors. Never forget this in the midst of your diagrams and equations »

Albert Einstein

Agradecimientos

En primer lugar quisiera agradecer a Dios y mi Familia (Amaya-Ramírez-Ruiz-Parra) por el amor, paciencia y apoyo incondicional que me han brindado. A ellos debo todo lo que soy.

Agradezco a mis directores de tesis, los profesores Luis Fernando Niño Vásquez y Carlos Alberto Parra Lopez por su consejo, guía y paciencia en este proceso.

Agradezco a los diferentes integrantes de los grupos de investigación *Inmunología y Medicina Traslacional* (I&MT) y *Laboratorio de Investigación en Sistemas Inteligentes* (LISI) por sus múltiples y valiosos aportes durante el desarrollo del presente trabajo.

Agradezco a los profesores Dora Bernal de Burgos, José Ismael Peña Reyes y Maria Alejandra Guzmán Pardo por brindarme su apoyo a través del Programa de Relaciones Internacionales de la Facultad de Ingeniería, apoyo sin el cual no hubiera podido desarrollar mis estudios.

Agradezco a la Universidad Nacional de Colombia (mi alma mater), a quien debo toda mi formación académica, por brindarme los medios necesarios para desarrollar la presente investigación.

Agradezco al profesor Andres Mauricio Pinzón Velasco por sus aportes al desarrollo de la presente tesis. También agradezco al Instituto de Genética por permitirme el uso de su infraestructura de cómputo.

Finalmente, agradezco a todos mis amigos que me acompañaron en este proceso por su ánimo y apoyo en todo momento.

Resumen

En el presente trabajo se expone la implementación de una estrategia *in-silico* para la identificación y priorización de péptidos candidatos a vacuna personalizada en tumores de cáncer aplicado a un caso de estudio de una paciente con PEComa que expresa el alelo HLA-A*24:02. Dicha estrategia se compone de 2 grandes etapas. La primera etapa consiste en la identificación de péptidos candidatos a vacuna personalizada a través de análisis de datos genéticos (ADN y ARN) que permiten la identificación y cuantificación de expresión de mutaciones somáticas presentes en el tumor, a partir de las cuales se obtienen todos los posibles péptidos mutantes (de entre 9 y 21 aminoácidos de longitud) que podría expresar el tumor; estos péptidos son filtrados a través de algoritmos que evalúan la afinidad y estabilidad con el haplotipo HLA del paciente para obtener una lista reducida de péptidos candidatos a vacuna personalizada. En la segunda etapa se realizan simulaciones de *docking* molecular entre los péptidos cortos previamente identificados (de entre 9 y 10 aminoácidos de longitud) y la molécula HLA-A*24:02 con el fin de evaluar y caracterizar la interacción péptido-HLA de tal manera que proporcione información que apoye el proceso de priorización de péptidos a evaluar de manera *in-vitro*.

Para nuestro caso de estudio, se identificaron 12 péptidos candidatos a vacuna personalizada (6 de los cuales son péptidos largos que enmarcan una epítipo tanto para moléculas HLA clase I como para clase II), de las cuales se simuló el *docking* molecular de 5 péptidos cortos (tanto la versión mutada como la nativa) con la molécula HLA-A*24:02. Dichas simulaciones permitieron identificar los puentes de hidrógeno entre el péptido y la molécula HLA y realizar una estimación de la energía del complejo, lo cual llevó a priorizar 2 de los 5 péptidos evaluados.

Aunque la estrategia permitió identificar y priorizar péptidos candidatos a vacunas personalizadas en un tumor de una paciente con PEComa, queda como perspectiva extender la estrategia del *docking* molecular a péptidos largos con moléculas HLA clase II y evaluar la estabilidad del complejo péptido-HLA a través de dinámica molecular.

Palabras clave: Vacunas personalizadas, cáncer, docking molecular

Abstract

In the present work we present the implementation of an *in-silico* strategy for the identification and prioritization of peptide candidates for personalized vaccine in cancer tumors applied to a case study of a patient with PEComa expressing the allele HLA-A*24:02. This strategy consists of two major stages. The first stage consists of the identification of peptide candidates to personalized vaccine through the analysis of genetic data (DNA and RNA) that allow the identification and quantification of expression of somatic mutations present in the tumor, from which all possible mutant peptides (between 9 and 21 amino acids in length) that could express the tumor are obtained; these peptides are filtered through algorithms that evaluate the affinity and stability with the HLA haplotype of the patient to obtain a reduced list of peptide candidates for a personalized vaccine. In the second stage, molecular docking simulations are performed between the previously identified short peptides (between 9 and 10 amino acids in length) and the HLA-A*24:02 molecule in order to evaluate and characterize the HLA-peptide interaction in such a way as to provide information that supports the process of prioritization of peptides to be evaluated *in-vitro*.

For our case study, 12 peptides were identified as candidates for personalized vaccine (6 of which are long peptides that frame an epitope for both HLA class I and class II molecules), of which the molecular docking of 5 short peptides (both the mutated and wildtype versions) were simulated with the HLA-A*24:02 molecule. These simulations allowed to identify the hydrogen bridges between the peptide and the HLA molecule and to estimate the energy of the complex, which led to prioritizing 2 of the 5 peptides evaluated.

Although the strategy allowed to identify and prioritize peptide candidates for personalized vaccines in a tumor of a patient with PEComa, it is possible to extend the molecular docking strategy to long peptides with HLA class II molecules and to evaluate the stability of the HLA-peptide complex through molecular dynamics.

Keywords: Personalized vaccines, cancer, molecular docking

Contenido

	Pág.
Introducción	11
Definición del problema	12
Objetivo general y objetivos específicos	13
Objetivo general	13
Objetivos específicos	13
Marco teórico	14
Incidencia del cáncer	14
Sobre el cáncer	14
Sobre el PEComa	15
Procesamiento y presentación de antígenos por el sistema inmunológico	18
Diseño de vacunas terapéuticas individualizadas para cáncer basadas en péptidos	21
Secuenciación de última generación	23
Identificación y anotación de variantes genómicas	25
Guías de buenas prácticas para la identificación y anotación de variantes	26
Identificación y priorización de neoantígenos candidatos a vacuna personalizada para cáncer	29
Generación de archivo FASTA conteniendo las secuencias de péptidos	30
Ejecución del software de predicción de las epitopes	31
Procesamiento y filtrado de los resultados	31
Modelamiento molecular	31
Acople molecular (docking)	32
Cálculo de la energía	34
Acoplamiento molecular de complejos péptido-proteína	35
Rosetta FlexPepDock	36
Descripción del protocolo Rosetta FlexPepDock	36
Materiales y métodos	39
Resumen de historia clínica de la paciente	39
Datos ómicos	39
Resultados y Discusión	43

Conclusiones y perspectivas	54
Conclusiones	54
Perspectivas	54
Bibliografía	56
Anexos	60
Anexo 1: Resultados del control de calidad realizado con FastQC	60
Anexo 2: Resultados tipificación haplotipo HLA	89
Anexo 3: Resultados detallados de la predicción de epítopes	90
Anexo 4: Resultados del docking molecular	91

Lista de figuras

	Pág.
Figura 2-1: Angiomiolipoma renal: células epitelioides perivasculares dispuestas alrededor de un vaso sanguíneo	15
Figura 2-2: Procesamiento y presentación de antígenos	19
Figura 2-3: Estructura de una molécula del MHCI	19
Figura 2-4: Metodología para la identificación de neoantígenos	23
Figura 2-5: Etapas del preprocesamiento de datos genéticos	28
Figura 2-6. Generación de secuencias de péptidos y filtrado de las epítopes candidatas predichas	30
Figura 2-7: Descripción general de los protocolos Rosetta FlexPepDock	37
Figura 3-1: Flujo de trabajo de la identificación y priorización de neoantígenos	41
Figura 3-2: Etapas del protocolo implementado para la ejecución del <i>docking</i> molecular	42
Figura 4-1: Puntajes promedio de calidad [Q_{Phred}] por posición en la lectura para el exoma normal (lecturas <i>Forward</i>)	43
Figura 4-2: Puntajes promedio de calidad [Q_{Phred}] por posición en la lectura para el exoma normal (lecturas <i>Reverse</i>)	44
Figura 4-3: Puntajes promedio de calidad [Q_{Phred}] por posición en la lectura para el exoma tumoral (lecturas <i>Forward</i>)	44
Figura 4-4: Puntajes promedio de calidad [Q_{Phred}] por posición en la lectura para el exoma tumoral (lecturas <i>Reverse</i>)	45
Figura 4-5: Puntajes promedio de calidad [Q_{Phred}] por posición en la lectura para el transcriptoma tumoral (lecturas <i>Forward</i>)	45
Figura 4-6: Puntajes promedio de calidad [Q_{Phred}] por posición en la lectura para el transcriptoma tumoral (lecturas <i>Reverse</i>)	46
Figura 4-7: Distribución de las variantes de acuerdo a su tipo	47
Figura 4-8: Distribución de las variantes de acuerdo al tipo de consecuencia en	

regiones codificantes	47
Figura 4-9: Distribución de variantes por cromosoma	48
Figura 4-10: Distribución de variantes por posición en proteína	48

Lista de tablas

	Pág.
Tabla 2-1: Clasificación de los tumores PE	17
Tabla 2-2: Localización de las moléculas del MHC	20
Tabla 3-1: Características de las secuenciaciones utilizadas para implementar y estandarizar la estrategia in-silico de identificación y priorización de epítopes potencialmente inmunogénicas	40
Tabla 4-1: Haplotipo HLA de la paciente	49
Tabla 4-2: Lista de péptidos candidatos a vacuna (MT: Mutant, WT: WildType)	50
Tabla 4-3: Lista de péptidos simulados (MT: Mutant, WT: WildType). En fondo azul los péptidos a los que se les realizó <i>docking</i> molecular con la molécula HLA-A*24:02	51
Tabla 4-4: Cálculo del RSMD entre el backbone del receptor resultante del docking y el backbone del receptor de la plantilla respectiva para todos los complejos simulados	51
Tabla 4-5: Identificación de puentes de hidrógeno en la interface péptido-HLA	52
Tabla 4-6: Valores de la función de puntuación de Rosetta obtenidos durante el <i>docking</i> molecular	52

1. Introducción

Gracias a la comprensión del funcionamiento del sistema inmune en el cáncer, en los últimos años se han desarrollado nuevas estrategias terapéuticas contra el cáncer denominadas “inmunoterapias” dentro de las cuales se encuentran las vacunas personalizadas. Este tipo de vacunas está dirigida contra neoantígenos tumorales generados a partir de mutaciones somáticas expresadas específicamente por el tumor, pero no por células normales. Los neoantígenos tumorales son reconocidos como extraños por linfocitos T excluidos del proceso de inducción de tolerancia en el timo, por lo anterior los linfocitos que los reconocen tienen la capacidad de inducir de manera eficiente una respuesta de rechazo contra el tumor que los expresan.

La identificación de neoantígenos puede hacerse hoy de manera personalizada en cada paciente en dos pasos: la secuenciación de los ácidos nucleicos extraídos de muestras del tumor (lo cual permite establecer el universo de neoantígenos del tumor), seguida del análisis bioinformático de los neoantígenos identificados. El análisis bioinformático (1, 2) permite analizar la información generada por una parte de la secuenciación del exoma (regiones codificantes del ADN), con la que se determinan y caracterizan las mutaciones codificantes para los neoantígenos producidos por el tumor; y por la otra, de la secuenciación del ARN tumoral con la cual es posible determinar el nivel de transcripción de los fragmentos del ARN que codifican para los neoantígenos. Una vez establecido el universo de mutaciones y su nivel de transcripción se procede a tamizar mediante análisis *in-silico* el grado de estabilidad de complejos formados por moléculas MHC expresadas por el tumor y péptidos de entre 8 y 10 aminoácidos que comprenden una mutación del tumor (complejos MHC-péptido). De la estabilidad de un complejo MHC-péptido depende en buena medida la inmunogenicidad de un péptido la cual se define como la capacidad del péptido de formar complejos MHC-péptido estables que estimulen con eficiencia el receptor para el antígeno en células T antitumorales (TCR por sus siglas en inglés).

Este tipo de vacuna terapéutica representa hoy una alternativa viable y esperanzadora para algunos tipos de cáncer de difícil tratamiento como es el caso del cáncer de páncreas, glioblastoma, carcinoma de mama triple negativo o el PEComa, para los cuales los protocolos de tratamiento convencionales (extirpación tumoral, uso de radiación ionizante y utilización de químicos

citostáticos y citotóxicos para eliminar e inhibir el crecimiento de células cancerosas) con frecuencia fallan.

1.1 Definición del problema

Para que un péptido (neoantígenos) sea inmunogénico para un linfocito T antitumoral es necesario que el TCR de este linfocito sea estimulado eficientemente por un complejo MHC-péptido. El proceso de identificación de péptidos potencialmente inmunogénicos es extremadamente complejo debido a que el espacio de búsqueda puede llegar a ser muy grande dada la gran cantidad de combinaciones distintas que se pueden formar con los 20 aminoácidos naturales. Para el caso 'simple' de un epítipo de 8 aminoácidos de longitud (que es la longitud mínima requerida para que un péptido se una a una molécula MHC), se podrían generar 10^{10} posibles combinaciones de aminoácidos. A pesar de resultados recientes prometedores de algunas investigaciones en materia de diseño de vacunas personalizadas útiles para la inmunoterapia del cáncer (3–9) la tasa de falsos positivos en el proceso de identificación de epítopos inmunogénicos sigue siendo alta, debido a que la predicción de la unión de péptidos a moléculas MHC se hace con base en algoritmos predictivos *in-silico* aún con limitada precisión y exactitud (10–12) quizás en parte debido a que aspectos moleculares del procesamiento y presentación del antígeno por las células presentadoras del antígeno; aspectos de la estabilidad del complejo MHC-péptido y del reconocimiento del complejo MHC-péptido por el TCR del linfocito T (13, 14) - factores todos determinantes de la inmunogenicidad del antígeno para los linfocitos T - aún no han podido ser incorporados a los algoritmos existentes.

En el presente trabajo se implementó por primera vez en Colombia una metodología *in-silico* que permitió identificar y priorizar péptidos potencialmente inmunogénicos útiles para la inmunoterapia del cáncer, utilizando como modelo la información genética del exoma de tejidos normal y tumoral con el transcriptoma de tejido tumoral, en el caso de un PEComa (tumor altamente maligno de células epitelioides perivasculares). Utilizando *docking* (acople) molecular la información estructural de complejos formados entre la molécula HLA-A*24:02 y neoantígenos seleccionados mediante métodos *in-silico*, fue obtenida y cotejada con el fin de explorar la utilidad del análisis estructural en la selección de neoantígenos.

1.2 Objetivo general y objetivos específicos

1.2.1 Objetivo general

Diseñar e implementar una estrategia *in-silico* que permita identificar y priorizar epítopes potencialmente inmunogénicas en cáncer.

1.2.2 Objetivos específicos

- 1) Identificar segmentos de péptidos mutantes candidatos a vacuna personalizada en el tumor de una paciente con PEComa que expresa el alelo HLA-A*24:02.
- 2) Utilizar *docking* molecular para identificar características que puedan favorecer la inmunogenicidad de péptidos candidatos a vacuna personalizada en un tumor de una paciente con PEComa que expresa el alelo HLA-A*24:02.

2. Marco teórico

2.1 Incidencia del cáncer

A nivel mundial, según cifras del Global Cancer Observatory (GCO), para 2018 el cáncer afectó a cerca de 18 millones de personas y proyecta que en el año 2040 habrá 29.5 millones de personas con esta afección (15). Respecto a su mortalidad, el cáncer causó aproximadamente 9 millones de casos en el 2016, siendo así una de las principales causas de muerte entre las enfermedades no transmisibles (16).

En Colombia, según cifras del GCO, para 2018 se estimaron 101.893 nuevos casos de personas con cáncer, 46.057 muertes asociadas a esta enfermedad y una prevalencia (para 5 años) de 230.726 casos (17). Entre los hombres las principales localizaciones fueron próstata, estómago, pulmón, colon-recto y ano, y linfomas no Hodgkin. En mujeres, las principales localizaciones fueron mama, cuello del útero, colon-recto y ano, estómago y tiroides.

Estas cifras ofrecen un panorama del enorme impacto que tiene esta enfermedad en la sociedad. Por lo cual, se hace necesaria la investigación en esta área, con el fin de desarrollar nuevos tratamientos que reduzcan los índices de mortalidad de esta enfermedad.

2.2 Sobre el cáncer

El cáncer es una enfermedad causada por una serie de mutaciones en genes las cuales son capaces de alterar la función celular (18). Dichas alteraciones afectan el ciclo celular, conduciendo a una proliferación anormal. Más particularmente son los proto-oncogenes, involucrados en la división y crecimiento celular, los que se ven afectados por las mutaciones cambiando así a oncogenes. Además, para que las células cancerígenas sean capaces de proliferar y perdurar, es necesario que se produzcan fallos tanto en los mecanismos intracelulares de reparación del ADN, como en los mecanismos extracelulares de vigilancia del sistema inmune (19-21).

Respecto a esto último, gracias a los avances en la comprensión de las funciones del sistema inmune en el cáncer, se están desarrollando nuevas e innovadoras estrategias de inmunoterapia dentro de las cuales se encuentran la transferencia celular adoptiva (ACT por sus siglas en inglés) y

las vacunas terapéuticas personalizadas. La inmunoterapia del cáncer con vacunas personalizadas es una estrategia novedosa que muestran resultados favorables en algunos pacientes (9). Este tipo de vacunas está dirigida contra neoantígenos tumorales generados a partir de mutaciones somáticas expresadas específicamente por el tumor, pero no por células normales. Los neoantígenos tumorales son reconocidos como extraños por linfocitos T excluidos del proceso de inducción de tolerancia en el timo, por lo anterior los linfocitos que los reconocen tienen la capacidad de inducir de manera eficiente una respuesta de rechazo contra el tumor que los expresan.

Este tipo de estrategias terapéuticas se presentan como una alternativa viable y esperanzadora para algunos tipos de cáncer para los cuales los protocolos de tratamiento convencionales (extirpación tumoral, uso de radiación ionizante y utilización de químicos citostáticos y citotóxicos para eliminar e inhibir el crecimiento de células cancerosas) carecen de efectividad, como por ejemplo en el caso del PEComa (*Perivascular Epithelioid Cell tumour*).

2.3 Sobre el PEComa

Las células epitelioides perivasculares (PEC) (ver **Figura 2-1**) son un tipo de célula presente en un grupo de tumores que incluyen angiomiolipoma (tumores en tejido adiposo y tejido muscular que, por lo general, se encuentran en el riñón), tumor de “células claras” (CCST -*Clear-Cell “Sugar” Tumor*-) de pulmón y sitios extrapulmonares, linfangioleiomiomatosis (tumores en células musculares lisas de los pulmones, vasos sanguíneos y linfáticos pulmonares y las pleuras), tumor de células miomelanocíticas del ligamento falciforme y tumores de “células claras” raros de otros sitios anatómicos (22).

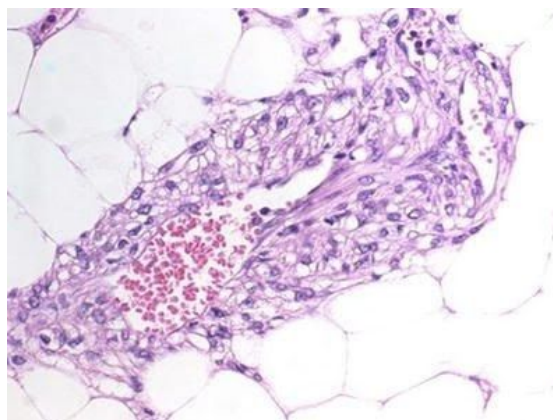


Figura 2-1: Angiomiolipoma renal: células epitelioides perivasculares dispuestas alrededor de un vaso sanguíneo. Imagen tomada de (22).

Los tumores de células epitelioides perivasculares (PEC por sus sigla en inglés) forman un grupo complejo de neoplasias que comparten la presencia de un tipo de célula único de origen y linaje aún desconocidos. Los tumores PEC o PEComas se desarrollan en numerosos órganos, incluidos el tracto hepatobiliar, el pulmón y el riñón. Las PEC se describieron por primera vez en 1944 por Apitz en su trabajo sobre tumores y malformaciones de la corteza renal (23). La descendencia neoplásica de las PEC comprende una familia de neoplasias, denominadas tumores PEC (**Tabla 2-1**). La Organización Mundial de la Salud (OMS) define los tumores de PEC o PEComas como tumores mesenquimatosos (tumores en tejido conjuntivo, los vasos sanguíneos y el tejido linfático) compuestos de células epitelioides perivasculares histológicamente e inmunohistoquímicamente distintivas. Las PEC como un linaje neoplásico en angiomiolipomas y tumores de células claras se observó por primera vez en 1991 (24). En 1992, se avanzó en la definición de una familia de tumores PEC, apoyándose en la detección de enlaces celulares entre angiomiolipomas, tumores de células claras y linfangioleiomiomatosis, y en que las células de estos tumores eran reactivas con marcadores de melanocitos, tenían una morfología epitelioides y una distribución perivascular (25). Los tumores de PEC muestran un amplio rango de comportamiento biológico, la mayoría de las lesiones son benignas o de comportamiento intermedio, pero también existen formas malignas (26). Algunos tumores son predominantemente epitelioides, mientras que otros se componen casi completamente de células grasas y, por lo tanto, se asemejan a los lipomas. Los tumores hepáticos de PEC muy raros incluyen tumores miomelanocíticos de células claras, otros tumores de células claras y linfangioleiomiomatosis (27).

Clasificación de los tumores PEC
Tumor de células epitelioides perivasculares sensu stricto (PEComa)
Tumor miomelanocítico de células claras (CCMT - <i>Clear Cell Myomelanocytic Tumor</i> -)
Tumor de células claras (CCST - <i>Clear-Cell "Sugar" Tumor</i> -)
Angiomiolipoma (AML)
Angiomiolipoma epitelioides monotípico (REON - <i>Renal Epithelioid Oxyphilic Neoplasm</i> -)
Microhamartoma Renal
Neoplasia renal capsular (RCN - <i>Renal Capsular Neoplasms</i> -)
Neoplasia renal epitelioides pediátrica con reactividad HMB-45
Nódulos de células claras pulmonares (CCN - <i>Clear Cell Nodules</i> -)
Linfangioleiomiomatosis (LAM)
Leiomioblastoma palisante intranodal
Sarcoma abdominopélvico de células epitelioides perivasculares

Tabla 2-1: Clasificación de los tumores PEC (27)

Entre las características morfológicas de las PEC se resalta su apariencia epitelioides con un citoplasma claro a granular, un núcleo redondo a ovalado, localizado centralmente y un nucleolo poco visible. Las PEC tiene una atipia leve y una localización perivascular típica (28). Inmunohistoquímicamente, el fenotipo prominente es el de una coexpresión de marcadores de células musculares y células melanóticas, principalmente HMB-45, Melan-A / MART-1, tirosinasa, factor de transcripción de microftalmia y actina del músculo liso (27). Las PEC generalmente no son reactivas para la proteína S-100, las citoqueratinas y la vimentina. A diferencia de los melanomas y los sarcomas de células claras, las PEComas son reactivas para MUM-1, un marcador de linfocitos conocido (28).

2.4 Procesamiento y presentación de antígenos por el sistema inmunológico

El sistema inmune dispone de dos mecanismos generales para generar una respuesta inmune, uno de ellos es a través de los linfocitos T. En esta vía de respuesta inmune, los linfocitos T reconocen un antígeno solamente cuando es presentado sobre la superficie de una célula

presentadora de antígeno en el marco de una molécula del complejo mayor de histocompatibilidad (MHC, por sus siglas en inglés) (29).

De manera breve, el MHC es un conjunto de moléculas (proteínas de membrana), las cuales determinan el repertorio de antígenos a los cuales los linfocitos T del individuo pueden responder, es decir, determinan en gran medida la respuesta inmune del individuo a antígenos propios y no propios. Por lo tanto, el MHC es responsable de la susceptibilidad o resistencia del individuo a agentes infecciosos, al desarrollo de fenómenos auto-inmunes y al desarrollo del cáncer. Existen dos clases de MHC, denominadas Clase I y Clase II (29).

En el presente estudio, el foco estará en las moléculas del MHC Clase I (MHCI). Estas moléculas se encargan de detectar antígenos endógenos (los producidos dentro de la célula, por ejemplo, antígenos de virus, algunas bacterias como la listeria, algunos parásitos como el *Tripanozoma cruzi* o neoantígenos en el caso del cáncer), los cuales en general son proteínas. Estas proteínas son direccionadas al proteasoma en donde son fragmentadas en péptidos. Estos péptidos son translocados al lumen del retículo endoplásmico por la molécula TAP (Transportador Asociado al Procesamiento en clase I), en donde son captados por las moléculas del MHCI. En caso de que la célula esté infectada por un virus o transformada por un tumor, esta célula presentadora del antígeno expone sobre la superficie celular un complejo MHCI-péptido el cual es detectado por linfocitos T CD8+, cuyo receptor es específico para el complejo MHCI-péptido que es presentado. Al reconocer este complejo, el linfocito T causa la muerte de la célula (29) (ver **Figura 2-2**).

Las moléculas MHCI son proteínas de membrana conformadas por una cadena Alfa que es polimórfica y la Beta-2 microglobulina que es monomórfica o conservada entre todos los individuos. La cadena Alfa tiene tres dominios Alfa-1, Alfa-2 y Alfa-3. La cadena Alfa se une a la membrana celular por el dominio transmembranal (dominio Alfa-3). La hendidura de unión de péptidos a la molécula MHCI está formado por el dominio Alfa-1 y Alfa-2 que son los más polimórficos de los tres dominios de la cadena Alfa. La hendidura tiene paredes formadas por hélices alfa y el piso por hojas plegadas tipo beta. La hendidura es estrecha en los extremos por lo que la longitud de los péptidos unidos es de 9 a 10 aminoácidos. En el piso de la hendidura se han identificado 6 bolsillos. En estos bolsillos se anclan las cadenas laterales de los aminoácidos del péptido (30) (**Figura 2-3**).

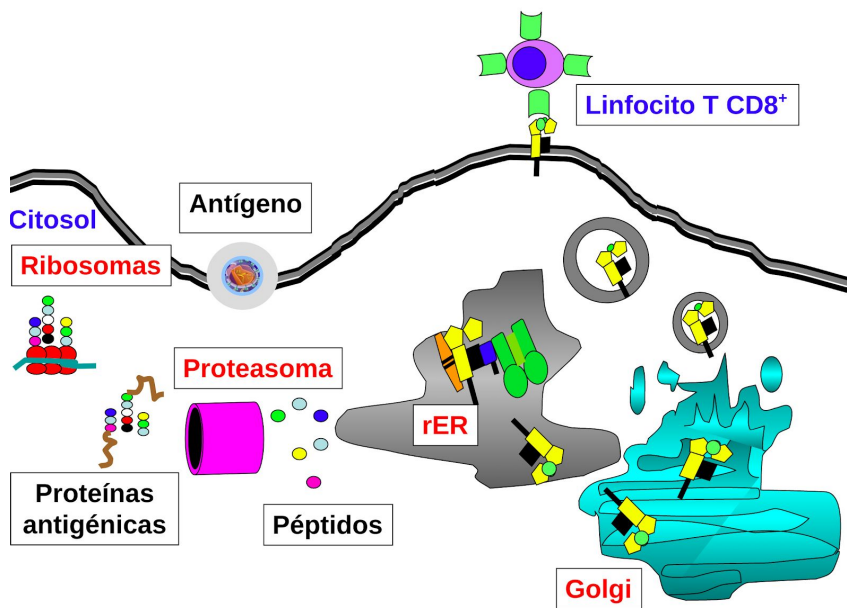


Figura 2-2: Procesamiento y presentación de antígenos. Imagen tomada y modificada de (31)

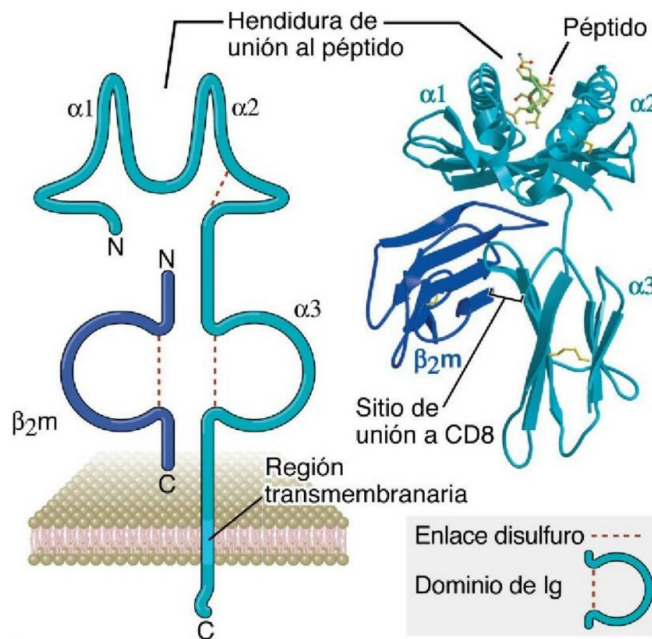


Figura 2-3: Estructura de una molécula del MHC I. Imagen tomada de (29).

Las moléculas del MHC I se expresan en todas las células nucleadas del organismo, a diferencia de las moléculas clase II, las cuales solo se expresan en células presentadoras del antígeno (APC, por sus siglas en inglés) profesionales tales como linfocitos B, macrófagos y células dendríticas (30) (ver **Tabla 2-2**).

Localización de las moléculas del MHC		
Tejido	Clase I	Clase II
Tejido Linfoide		
Células T	+++	+
Células B	+++	+++
Macrófagos	+++	++
Otras APC (Células dendríticas)	+++	+++
Células del epitelio tímico	+	+++
Otras células nucleadas		
Neutrófilos	+++	-
Hepatocitos	+	-
Riñón	+	-
Pulmón	+	-
Células no nucleadas		
Glóbulos Rojos	-	-

Tabla 2-2: Localización de las moléculas del MHC (30).

La activación de los linfocitos T por parte de las APC requiere de dos señales (30):

- Señal 1: Una vez ha procesado el antígeno, la célula presentadora del antígeno expresa un complejo MHCI-péptido o MHCII-péptido sobre su superficie. Este complejo es reconocido por el receptor para el antígeno de los linfocitos T (TCR, por sus siglas en inglés) el cual es específico para el respectivo complejo MHC-péptido (si el complejo es clase I-péptido este complejo es reconocido por el TCR de un linfocito T CD8+; si el complejo es clase II - péptido este es reconocido por un linfocito T CD4+ helper).
- Señal 2: El efecto de la señal 1 en el linfocito T produce como respuesta en esta célula la expresión de la molécula CD40 Ligando (CD40L) sobre su superficie. La expresión de CD40L en un linfocito T es un indicador de que ese linfocito T ha recibido la señal 1 y que este se ha activado. El CD40L interactúa entonces con un receptor sobre la célula presentadora del antígeno: la molécula CD40 (la cual es expresada de manera constitutiva por la célula presentadora del antígeno). La estimulación del CD40 por el CD40L produce en la célula presentadora del antígeno dos efectos:
 - 1) Expresión de la molécula de co-estimulación CD80 o CD86 que brindan la señal 2 al linfocito T que es la estimulación de la molécula CD28 sobre el linfocito T con lo cual el linfocito produce interleuquina-2 (IL-2) que es el factor de crecimiento de los linfocitos T por lo cual este comienza a proliferar.
 - 2) Si la célula presentadora del antígeno es una célula dendrítica o un macrófago, en respuesta a la estimulación de CD40 por el CD40L del linfocito T, la célula presentadora produce IL-12. La IL-12 es uno de los factores más importantes para la

activación de linfocitos T CD4+ helper tipo 1 (Th-1) y linfocitos T CD8+, ambos productores de IFN- γ (interferón gama). En respuesta a la producción de IFN- γ por el linfocito Th-1, la célula presentadora activa mecanismos microbicidas con lo cual mejora su capacidad para controlar la replicación intracelular de micro-organismos que esta ha fagocitado.

2.5 Diseño de vacunas terapéuticas individualizadas para cáncer basadas en péptidos

A pesar del considerable éxito de las terapias de bloqueo del control (*checkpoint blockade therapies*) en el ámbito clínico, las vacunas personalizadas se presentan como una necesidad clínica para aquellos pacientes que no responden a estas terapias, desarrollan resistencia o no pueden tolerar los efectos secundarios de las drogas utilizadas en dichas terapias. El paradigma principal detrás del desarrollo de vacunas para el cáncer reside en la hipótesis de que el sistema inmune puede ser estimulado para reconocer neoantígenos y, de esta manera, desencadenar una destrucción selectiva de las células tumorales. Este proceso dependerá de la capacidad del péptido/vacuna de interactuar con una molécula del MHC y ser presentado en membrana para eventualmente ser identificado por un linfocito T (1).

Varios estudios (3-6, 8, 9) han demostrado la viabilidad de desarrollar vacunas terapéuticas individualizadas para el tratamiento del cáncer. El principio de funcionamiento de esta novedosa aproximación experimental se basa en la identificación de segmentos de péptidos mutantes codificados por las células tumorales, los cuales presentan propiedades inmunogénicas. El complejo formado por la unión de dichos péptidos a moléculas del MHC I podría ser identificado por un linfocito T CD8+ del paciente (8). Estos antígenos únicos surgen como producto de los numerosos cambios genéticos que se producen en un proceso cancerígeno, dichos cambios genéticos son mutaciones somáticas que están presentes únicamente en el tumor, de ahí que se conozcan estos péptidos como péptidos mutantes.

Recientes estudios preclínicos indican que estas proteínas mutantes, pueden ser procesadas y presentadas en el marco de las moléculas del MHC y ser reconocidas como antígenos extraños por el sistema inmune (6).

El proceso de identificación de epitopes potencialmente inmunogénicas es extremadamente complejo, por ejemplo, para el caso 'simple' de un epítipo de 8 aminoácidos de longitud se podrían generar 10^{10} posibles combinaciones con los 20 aminoácidos naturales. Si se considera el caso de un epítipo de 16 aminoácidos de longitud, el número de combinaciones aumentaría a 10^{20} . Por lo cual,

se hace necesario diseñar una estrategia que permita identificar en un tiempo razonable las epítopes probables generadas en nuestro caso por un tumor. Por tal motivo, una estrategia *in-silico* se presenta como la opción más viable. Dicha estrategia *in-silico* se deberá apoyar en la información ofrecida por las técnicas de secuenciación de última generación de ADN y ARN, pues a partir de esta información se podrán identificar las mutaciones no sinónimas traducidas a proteína. A partir de esta información, estudios recientes han planteado la siguiente metodología: realizar una predicción de epítopes, integrar la información derivada de los datos de secuenciación y luego seleccionar los posibles candidatos a ser neoantígenos (1). En la **Figura 2-4** se presenta un flujograma detallado de la metodología propuesta en (1).

A pesar de los resultados alentadores obtenidos en (1), la tasa de falsos positivos sigue siendo alta; esto se puede explicar en parte a que esta metodología se centra en el proceso de unión del péptido con la molécula del MHC. Sin embargo, otros factores pueden influenciar el proceso de inmunogenicidad, tales como el procesamiento del antígeno por parte de la célula (32) o la estabilidad del complejo péptido-molécula del MHC (13).

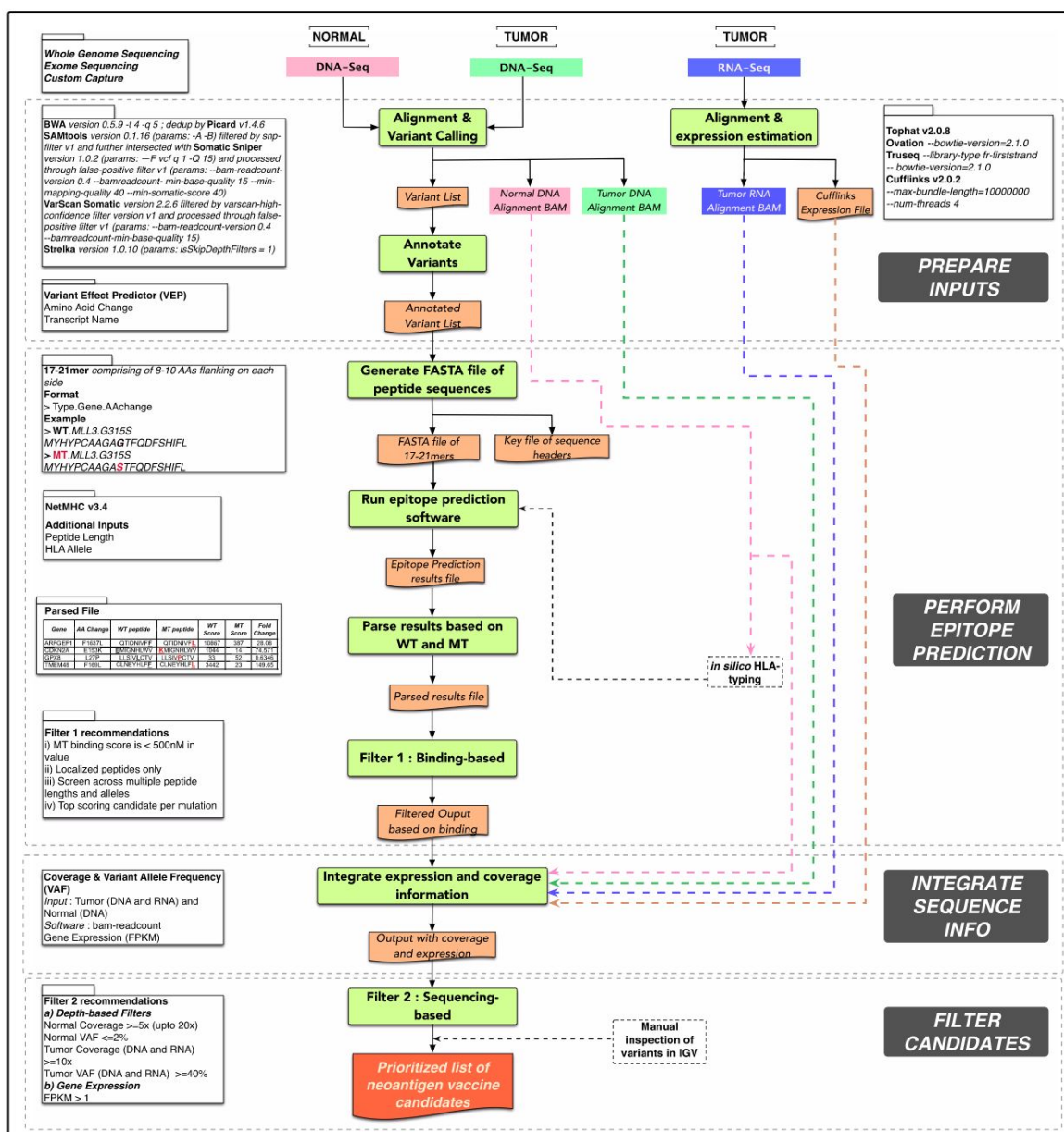


Figura 2-4: Metodología para la identificación de neoantígenos. Imagen tomada de (1).

2.5.1 Secuenciación de última generación

La secuenciación de nucleótidos (ADN o ARN) es un conjunto de métodos y técnicas bioquímicas cuya finalidad es la determinación del orden de los nucleótidos (A, C, G, T o U) en un oligonucleótido de ADN o ARN. La secuenciación de última generación (NGS, por sus siglas en inglés) abarca un conjunto de técnicas de secuenciación masivas y en paralelo que producen enormes cantidades de lecturas (segmentos de nucleótidos) de alta calidad, de cortos tamaños (entre 70 y 200 pares de bases -pb-), a alta velocidad y a un costo razonable, permitiendo de esta

manera la investigación básica en genómica, transcriptómica, metagenómica o proteómica, entre otras ómicas (33).

Desde la introducción de las tecnologías NGS en 2007, el precio de la secuenciación de todo un genoma humano se ha reducido rápidamente. A principios de 2014, Illumina Inc. introdujo la plataforma HiSeq X en el mercado, reduciendo el precio de un genoma personal a 1.000 USD (costo experimental). Con esta nueva máquina, se pueden secuenciar 16 genomas humanos en 3 días a una profundidad de 30× (34). Se han establecido varias iniciativas nuevas para aprovechar esta dramática caída de precios para secuenciar miles de genomas humanos. Por ejemplo, el proyecto *Genome England* tiene por objetivo secuenciar 100,000 genomas y los adjuntará a los datos de los registros médicos como parte del sistema médico del Reino Unido. *Human Longevity Inc.*, fundada por J. Craig Venter y otros dos socios, anunció que construirá el centro de secuenciación del genoma humano más grande del mundo con la capacidad de secuenciar hasta 40,000 genomas humanos por año. Se espera que la caída en el precio de la secuenciación por base (nucleótido) impulse la generación de una inmensa cantidad de datos NGS, creando un desafío de *Big Data* en bioinformática.

Las secuenciaciones tipo NGS producen de millones a miles de millones de lecturas de secuencias cortas a alta velocidad. Las plataformas NGS actuales que incluyen Illumina, Ion Torrent de Life Technologies, Pacific Bioscience, Nanopore y GenapSys pueden generar lecturas de 100–10,000 bases de largo (35–37) permitiendo una mejor cobertura del genoma a un menor costo. Sin embargo, estas plataformas también generan enormes cantidades de datos en bruto. Por ejemplo, los datos sin procesar producidos por la plataforma Illumina HiSeq2500 suman hasta 1 TB por corrida, cubriendo 150–180 muestras de secuenciación de exomas humanos completos (WES por sus siglas en inglés) a una profundidad de 50× o más (Illumina Inc.). Para la identificación de variantes en cáncer se recomienda secuenciar el exoma normal a una profundidad de aproximadamente 150× y el exoma tumoral a una profundidad de aproximadamente 300× con el fin de garantizar una sensibilidad y confiabilidad suficientes en la identificación de las variantes y en el caso del exoma tumoral compensar la posible “contaminación” de la muestra con ADN de células normales (2).

En este trabajo el objetivo de la secuenciación de ADN es descubrir variantes genómicas en forma de variantes de un solo nucleótido (SNV por sus siglas en inglés), pequeñas inserciones o eliminaciones de ADN (InDels, abreviación para *insertions* y *deletions* en inglés), variaciones en el número de copias (CNV por sus siglas en inglés) u otras variantes estructurales (SV por sus siglas en inglés), con el objetivo final de asociar esas variantes a enfermedades humanas. La secuencia de

ARN que mide los cambios en la expresión génica se puede usar para descubrir nuevos transcritos, incluidos los ARN no codificantes y detectar el empalme de transcritos o los eventos de fusión de genes.

Sin embargo, este tipo de técnicas de secuenciación plantean algunos retos dada la gran cantidad de datos que generan, entre ellos se destaca el ensamblaje de las lecturas cortas. Esto debido al hecho de que estas técnicas se basan en un método de secuenciación aleatorio, el cual consiste en romper el genoma al azar y secuenciar segmentos cortos de manera independiente. Lo anterior implica la necesidad de realizar un proceso de reconstrucción del genoma a través de la unión de las lecturas hasta llegar a reconstruir cada uno de los cromosomas, a este proceso se le conoce como ensamblaje del genoma (33).

Existen dos tipos de ensamblajes: ensamblaje *de novo* y ensamblaje por referencia. El primero consiste en unir las lecturas de tal manera que se formen secuencias contiguas (*contigs*) que compartan las mismas secuencias de nucleótidos de acuerdo al ADN plantilla de donde provenían las secuencias, luego se unen dichos *contigs* con el fin de obtener secuencias más largas llamadas *scaffolds* a partir de las cuales se reconstruyen los cromosomas. El segundo, por otro lado, utiliza un genoma de referencia (un genoma ya secuenciado y conocido de la misma especie) respecto al cual se mapean las lecturas y, posteriormente, se reconstruye el genoma (38).

2.5.2 Identificación y anotación de variantes genómicas

El advenimiento de las tecnologías de secuenciación de última generación ha promovido en gran medida avances en el estudio de las enfermedades humanas en los niveles genómico, transcriptómico y epigenético. La secuenciación del exoma, donde la región codificante del genoma se captura y se secuencia, ha demostrado ser un método rentable para detectar variantes causantes de enfermedades y descubrir marcadores genéticos.

Para el caso del presente estudio, se tiene por interés identificar las mutaciones únicas que presentan las células tumorales respecto a las células normales. Con esto en mente, se hace necesario diseñar un flujo de trabajo que permita realizar una identificación confiable. En este sentido, un flujo de trabajo típico para el análisis de WES consiste en los siguientes pasos: control de calidad de los datos sin procesar, preprocesamiento, alineamiento, procesamiento post-alineamiento, identificación de variantes, anotación y priorización (39).

2.5.2.1 Guías de buenas prácticas para la identificación y anotación de variantes

El *Broad Institute* es un referente a nivel mundial en lo que respecta a análisis genómicos. Esta institución nació en 2004 como una iniciativa de varios científicos del *Massachusetts Institute of Technology* (MIT) y de *Harvard University* con el objetivo de desarrollar investigaciones en el área de la genómica que derivaran en mejoras para la salud humana, ya sea avanzando en la comprensión de la biología de enfermedades humanas o desarrollando nuevos tratamientos para estas. Los principales ejes de investigación del *Broad Institute* son: enfermedades infecciosas, cáncer, psiquiatría y enfermedades cardiovasculares.

Entre los múltiples desarrollos que ha realizado el *Broad Institute* se resalta una *suite* de herramientas bioinformáticas llamada *Genome Analysis ToolKit* (GATK) las cuales se centran en la identificación de variantes y genotipificación en humanos. Adicionalmente, han desarrollado unas guías de referencia de “Buenas Prácticas” para múltiples análisis genómicos. Estas guías que ya van en su cuarta versión y los detalles de su versión original se encuentran en (39). Estas guías proveen recomendaciones paso a paso para, por ejemplo, realizar identificación de variantes a partir de datos de secuenciación de última generación, intentando describir en detalle los principios clave de los pasos de procesamiento y análisis necesarios para ir desde las lecturas “crudas” que salen de las máquinas de secuenciación, hasta un conjunto de variantes correctamente filtradas que se puedan usar en análisis posteriores.

En el caso particular de la identificación y anotación de variantes, este análisis se compone de tres grandes fases:

1. Pre-procesamiento de los datos: en esta fase la idea es procesar los datos de tal manera que se pase de los datos “crudos” (normalmente en formato FASTQ) a datos limpios, organizados y listos para analizar en formato BAM (*Binary Alignment Map*). Esto implica realizar alineamientos *de novo* o por referencia y limpieza para filtrar datos erróneos o sesgados. En la **Figura 2-1** se presentan los pasos a realizar en esta primera fase.

El primer paso consiste en realizar un análisis de calidad que permita identificar defectos o errores de la secuenciación y, de esta manera, tomar las medidas pertinentes para filtrar las lecturas o segmentos de lecturas que no cumplan con los criterios de calidad que se establezcan.

El segundo paso consiste en realizar el alineamiento de las lecturas para cada conjunto de secuenciaciones, es decir, cada lectura pareada (para el caso de una secuenciación *pair-end*)

es mapeada a una posición respecto a un genoma de referencia (esto para el caso de un alineamiento por referencia). Este genoma de referencia consiste en una representación a una hebra de un genoma definido como común para la especie del individuo al cual corresponde la secuenciación. Por lo tanto, funciona como un marco de coordenadas de referencia comunes para todos los análisis genómicos.

El tercer paso consiste en la identificación de lecturas pareadas con alta probabilidad de ser duplicados del mismo fragmento original de ADN debido a algún “artefacto” del proceso de secuenciación. Este tipo de lecturas se consideran observaciones no independientes, por lo que se busca ya sea etiquetar o eliminar todas estas lecturas excepto una con el fin de que no influyan en el proceso de identificación de variantes. Luego, se hace necesario ordenar las lecturas respecto a sus coordenadas en el genoma de referencia.

Finalmente, el cuarto paso consiste en realizar una recalibración de los puntajes de calidad por base secuenciada. Para ello se recurre a herramientas de aprendizaje de máquina (en inglés *machine learning*) para detectar y corregir patrones de errores sistemáticos en los puntajes de calidad por base. Estos puntajes son un indicador de la certeza con que el secuenciador identifica una base. La importancia de estos puntajes yace en que permiten ponderar la evidencia en pro o contra de posibles variantes. De aquí la importancia de corregir cualquier sesgo sistemático introducido durante la secuenciación. Dichos sesgos pueden provenir de los procesos bioquímicos utilizados en la preparación de la librería o en la secuenciación, los defectos de fabricación en los chips o la instrumentación del secuenciador. El proceso de recalibración involucra el cálculo de valores estadísticos de covarianza sobre la totalidad de los datos, la construcción de un modelo a partir de dichas estadísticas y la aplicación del ajuste a los puntajes de calidad de acuerdo al modelo construido.

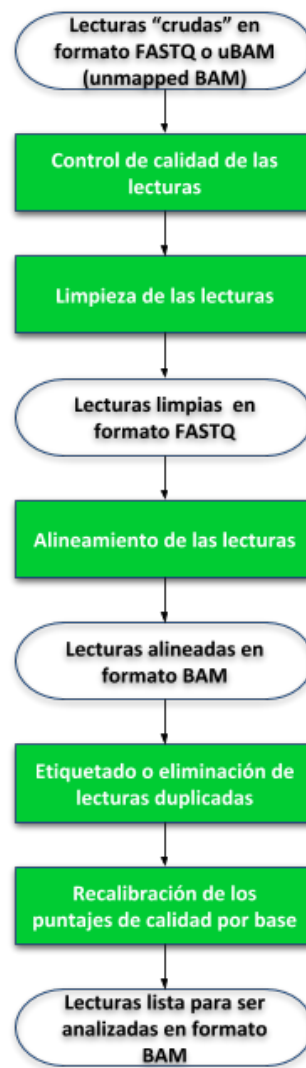


Figura 2-5: Etapas del preprocesamiento de datos genéticos.

2. Identificación de variantes: A partir de los datos en formato BAM de la fase anterior, se procede a identificar variantes genéticas a través de la comparación entre los diferentes conjuntos de datos, por ejemplo, para la presente investigación la comparación se haría entre los exomas normal y tumoral. Típicamente, los tipos de variantes más relevantes para este tipo de estudios son las variantes de un único nucleótido (SNV por sus siglas en inglés) y los indels (inserciones y deleciones). De manera general, la identificación de las variantes se realiza a través de ensamblajes *de-novo* locales en las regiones activas (regiones donde existen candidatos a variantes); es decir, se identifican regiones que presentan indicios de variantes (regiones en donde las lecturas mapeadas en regiones análogas de los alineamientos de los exomas normal y tumoral presentan discrepancias significativas), se re-ensamblan por completo dichas regiones utilizando el método de ensamblaje *de-novo*, se alinean las regiones análogas de los conjuntos de datos y se aplican modelos estadísticos

(usualmente modelos bayesianos) para determinar la probabilidad de que la discrepancia identificada sea una variante somática o sea producto de errores de la secuenciación.

El resultado de este proceso típicamente genera un archivo en formato VCF (*Variant Call File*).

3. Anotación de variantes: A partir de bases de datos de variantes conocidas, de funciones biológicas, de consecuencias genéticas, de asociaciones a enfermedades, etc, a cada una de las variantes previamente identificadas se pueden asociar funciones o características de tal manera que posteriormente se puedan realizar análisis más profundos o para el presente caso filtrar variantes. Algunos ejemplos de estas bases de datos son 1000 Genomes, Ensembl, RefSeq, GENCODE, dbSNP, gnomAD, COSMIC, etc.

2.5.3 Identificación y priorización de neoantígenos candidatos a vacuna personalizada para cáncer

El proceso de identificación de neoantígenos se apoya fundamentalmente en el proceso de unión y presentación de neoantígenos por parte de las moléculas del HLA, debido a que este es un paso crítico para inducir una respuesta inmune (40).

Actualmente existen múltiples herramientas *in-silico* que permiten estimar la afinidad de unión de un péptido con moléculas del HLA (10, 11, 41–43), el procesamiento de un péptido por el proteasoma (44) y la vida media de un complejo péptido-HLA (45). Adicionalmente existen varias herramientas *in-silico* que apoyadas en las anteriores, permiten realizar una identificación y priorización de neoantígenos. Una ejemplo es pVAC-seq (1), la cual es una de las pocas de código abierto y que, además, ha sido probada en un ensayo de “prueba de concepto” para la identificación de neoantígenos en pacientes con melanoma para su uso como vacunas basadas en células dendríticas (6), obteniendo resultados muy interesantes.

El punto de partida de pVAC-seq es una lista de variantes/mutaciones en formato vcf anotada a través de VEP (46) y el haplotipo de la persona (este se puede obtener ya sea por ensayos clínicos de genotipificación o por métodos *in-silico* a partir del exoma de tejido normal (47, 48)). Como fue mencionado previamente, una de los componentes claves de pVAC-seq es la predicción de epítopes resultantes de las mutaciones gracias al cálculo de su afinidad de unión contra las moléculas del HLA. Este proceso involucra los pasos que se detallan a continuación (ver **Figura 2-6**):

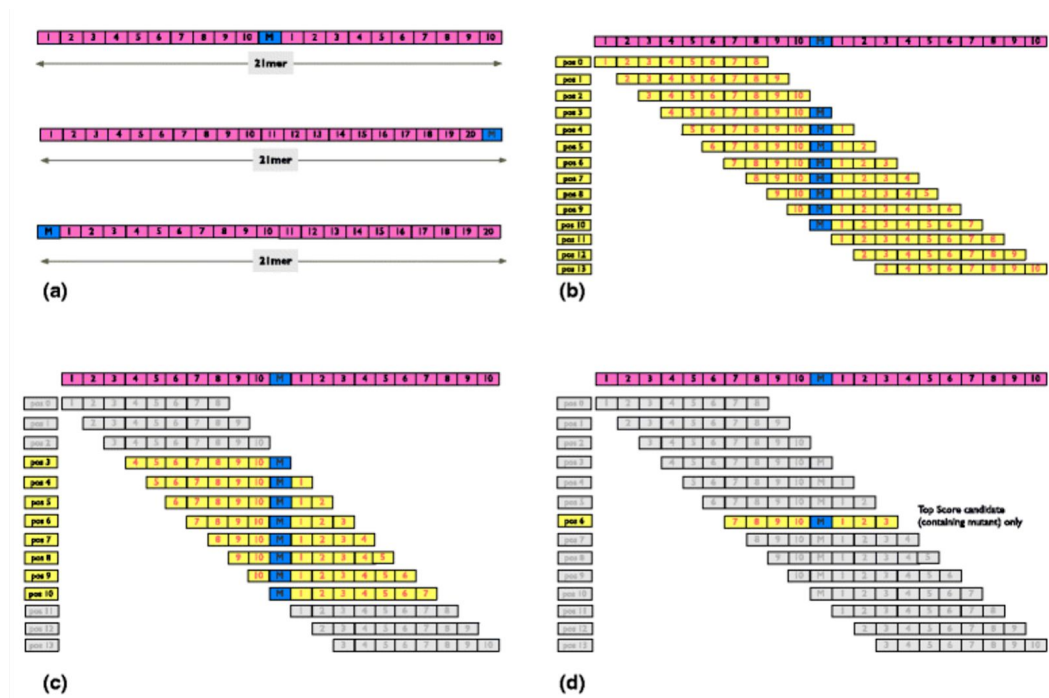


Figura 2-6. Generación de secuencias de péptidos y filtrado de las epítopes candidatas predichas. **a** Se construye una secuencia de aminoácidos en formato FASTA utilizando los 10 aminoácidos que flanquean cada lado de una mutación. En el caso de que la mutación se ubique cerca del final o inicio del transcrito, se toman los 20 aminoácidos previos o posteriores, respectivamente. **b** Se generan péptidos candidatos a través del mapeo con ventanas de tamaño dado por k-mers seleccionados con la secuencia previamente mencionada. **c** Sólo los segmentos de péptidos conteniendo la mutación o parte de ella (en el caso de indels o cambio de marco de lectura) son tenidos en cuenta. **d** Los mejores candidatos (los que presenten los puntajes más bajos de afinidad) por mutación y por molécula del HLA son retenidos. (Imagen tomada de (1)).

2.5.3.1 Generación de archivo FASTA conteniendo las secuencias de péptidos

Las secuencias de péptidos son uno de los elementos centrales para las herramientas que predicen la afinidad de unión (IC 50) con las moléculas del HLA. Con el fin de reducir el tiempo de cómputo, primero se genera un archivo FASTA que contenga 2 secuencias de péptidos por variante: WT (normal o Wild Type por sus siglas en inglés) y mutante (tumor). La secuencia FASTA se construye utilizando entre 8 y 10 aminoácidos que flanquean cada lado de la mutación. Sin embargo, si la mutación se ubica en las cercanías del fin o inicio del transcrito, se procede a considerar entre 16 a 20 aminoácidos previos o posteriores, respectivamente. Luego, se crea un

archivo con un encabezado (nombre y tipo de la variante) que permita posteriormente correlacionar cada epítipo con la información del encabezado.

2.5.3.2 Ejecución del software de predicción de las epitopes

pVAC-seq permite utilizar múltiples herramientas para la predicción de la afinidad (IC 50)péptido-HLA, entre las que se destacan NetMHC (42), NetMHCcons (43), NetMHCpan (10), NetMHCIpan (49), MHCflurry (11), PickPocket (41), SMMPMBEC (50), MHCnuggets (51), MHCnuggetsII (52) y SMMalign (53). Estas herramientas utilizan como información de entrada el haplotipo HLA del paciente y el archivo FASTA generado en la etapa anterior.

2.5.3.3 Procesamiento y filtrado de los resultados

Una vez se obtiene la lista de todas las posibles epítipes candidatas con los respectivos valores de afinidad péptido-HLA, se procede a aplicar ciertos filtros para seleccionar y priorizar los mejores candidatos. Primero, se seleccionan las epítipes candidatas que tengan puntajes de afinidad fuertes o moderados, es decir, que tengan puntajes de IC 50 inferiores a 500 nM. Segundo, se seleccionan las epítipes que contengan la mutación o parte de ella (casos de mutaciones derivadas de indels o de cambio del marco de lectura). Tercero, se selecciona el mejor candidato por mutación y por molécula del HLA. Cuarto, se seleccionan las epítipes de acuerdo con filtros basados en la profundidad de secuenciación y al VAF (Variant Allele Frequency). Quinto, integrando los datos de niveles de expresión de los transcritos que portan las epítipes candidatas, se seleccionan las epítipes que cumplan cierto umbral de expresión (usualmente 1 FPKM).

2.6 Modelamiento molecular

Dada la considerable cantidad y relevancia de procesos de señalización y regulación que involucran interacciones de un péptido con una proteína; en los últimos años y gracias a los enormes avances en capacidad de cómputo, han surgido múltiples herramientas de modelamiento molecular, los cuales permiten generar modelos precisos de estructuras péptido-receptor. Estos modelos ayudan a mejorar la caracterización y manipulación de interacciones conocidas, a descubrir interacciones péptido-proteína desconocidos e incluso a diseñar medicamentos basados en péptidos dirigidos a sistemas específicos de interés médico.

En general, dichas herramientas de modelamiento molecular se apoyan de la física y química para simular y caracterizar estructuras de moléculas o complejos de moléculas utilizando modelos teóricos tales como el de la mecánica clásica o el de la mecánica cuántica, en donde se consideran

parámetros como cargas, dipolos electrostáticos, radios de Van der Waals, parámetros de enlaces covalentes, ángulos diedrales, entre muchos otros.

Entre los principales beneficios de implementar análisis basados en modelamiento molecular se cuentan una mejor selectividad en los casos a estudiar experimentalmente y por ende menores tiempos y costos asociados a dichos estudios y reducción del trabajo a prueba y error.

Es importante hacer la aclaración de que este tipo de simulaciones no reemplazan los ensayos de laboratorio, sino que los complementan y nutren. Por ejemplo, los métodos de modelamiento molecular juegan un papel muy relevante en diversas ramas de la química, pues ayudan a interpretar resultados experimentales (54). Dependiendo de los análisis que se quieran realizar, estos métodos pueden requerir cierta información inicial que puede incluir la estructura del complejo receptor-ligando (en ciertos casos puede servir una estructura homóloga al complejo que se quiera simular), la estructura del receptor y la secuencia del ligando, sitios activos del receptor, etc. Estas estructuras moleculares pueden provenir de dos fuentes; la primera y más relevante fuente son los métodos experimentales como la cristalografía de rayos X o la resonancia magnética nuclear (la base de datos más representativa de este tipo de datos es PDB -*Protein Data Bank*- (55), la segunda fuente son los métodos *in-silico*, por ejemplo *I-TASSER* (56).

Uno de los análisis más frecuentes en el área del modelamiento molecular, es el de determinar la afinidad de un ligando con una región específica de un receptor. El interés en este tipo de análisis yace en que se pueden identificar ligandos que puedan inhibir o iniciar un proceso biológico, por ejemplo, un antígeno que al unirse y ser presentado por una molécula del HLA sea reconocido por un linfocito T, el cual inicie una respuesta inmune contra la fuente del antígeno (una bacteria, un virus, un microorganismo o células cancerígenas).

La investigación y el desarrollo de nuevas estrategias terapéuticas o profilácticas contra el cáncer, son procesos muy costosos en términos de tiempo y dinero. No obstante, el avance en modelos físicos y químicos para simular procesos biomoleculares, así como la mejora en la asequibilidad de recursos computacionales, abren la puerta al descubrimiento o desarrollo de nuevas terapias contra el cáncer.

2.6.1 Acople molecular (*docking*)

A través de la mecánica clásica se puede representar una molécula como un conjunto de átomos unidos entre sí por enlaces elásticos, los cuales se representan a su vez por resortes que pueden ser estirados, comprimidos o flexionados en diferentes ángulos. Por otro lado, la mecánica cuántica

permite modelar la estructura electrónica de una molécula a través de la ecuación de Schrödinger (57). A partir de esta ecuación, se puede calcular propiedades electrónicas de un sistema tales como la distribución de cargas en un átomo, interacciones electrostáticas, configuraciones estables de moléculas en reacción, etc.

En general, para estudios de conformación de complejos proteína-péptido básicos es suficiente la utilización de modelos basados en la mecánica clásica, ya que estos tienen un costo computacional muy inferior a aquellos basados en la mecánica cuántica y ofrecen una precisión aceptable.

El primer paso para realizar un acople molecular es identificar la información disponible sobre las moléculas a modelar, esto con el fin de identificar los datos de partida para el modelamiento y, de acuerdo a éstos últimos, seleccionar la herramienta apropiada. En particular, se busca identificar si existen estructuras 3D resueltas experimentalmente para el complejo o las moléculas individuales que se quieren modelar, y la localización de los sitios activos del receptor. Típicamente, para el caso del modelamiento de moléculas del HLA con péptidos, no se cuenta con la estructura 3D del complejo resuelta experimentalmente, en algunos casos se puede encontrar la estructura de la molécula del HLA unida a otro péptido y en el peor de los casos de ninguna de las moléculas. En el caso en el que se cuente con la estructura de la molécula del HLA con un péptido diferente al péptido de interés, existen herramientas tales como Pymol (58) o Chimera (59) que permiten reemplazar el péptido por el péptido de interés y a partir de esta estructura, se realiza el modelamiento del acople molecular. En el caso en el que no se cuente con ninguna estructura de las moléculas a modelar, existen herramientas *in-silico* como la ya mencionada I-TASSER las cuales predicen la estructura 3D de una proteína a partir de su secuencia de aminoácidos.

El proceso del modelamiento de un acople molecular se compone de dos grandes etapas: búsqueda y evaluación. En la primera se realiza una exploración del espacio conformacional accesible para el ligando sobre el receptor, en donde el objetivo es buscar la conformación que corresponda al mínimo global de energía libre de unión. En la segunda etapa se calcula una función de puntuación con la que se evalúa la afinidad de la interacción.

Estos métodos computacionales buscan encontrar sitios de interacción tanto geométricos como energéticos. Sin embargo, se pueden presentar varios problemas al momento de realizar el análisis y no obtener resultados adecuados, como son inexactitudes en el modelo de energía utilizados para dar la puntuación del complejo ligando/receptor y la incapacidad para determinar los cambios conformacionales que se producen durante el proceso de unión tanto para el ligando como para el receptor. Si el sistema fuera completamente rígido se podría hablar de una teoría de unión llave y

cerradura, donde entre la proteína y el ligando solo existe una forma en la que se podría formar el complejo. Sin embargo, hay diversas posibilidades de unión entre la proteína y el ligando debido a que se pueden realizar procesos de acoplamiento flexibles o semiflexibles, para lo cual es necesario realizar cálculos de energía que permitan determinar la conformación adecuada entre las dos moléculas. Los estados de energía que son evaluados durante el proceso de acoplamiento, determinan si la conformación establecida es óptima o si el algoritmo de búsqueda debe continuarse ejecutando hasta lograr obtenerse una conformación estable del sistema.

2.6.1.1 Cálculo de la energía

Las interacciones ligando-receptor pueden ser descritas a través de funciones de energía las cuales representen tres componentes de movimiento:

1. Movimiento de traslación y rotación del ligando.
2. Orientación del ligando respecto al receptor.
3. Ángulos diedros (ángulos de torsión).

Cada método computacional de modelamiento molecular tiene su estrategia de búsqueda del espacio conformacional entre el ligando y el receptor, así como el cálculo que permite evaluar cuál es la mejor configuración. Existen algunas herramientas que describen el sistema por medio de ecuaciones que consideran separadamente los principales elementos que contribuyen al cálculo de la energía, es decir, la energía asociada a interacciones intermoleculares e intramoleculares (60).

La energía asociada a interacciones intramoleculares es calculada entre el ligando y el receptor a través de la diferencia de los estados de enlace y no enlace de las dos moléculas:

$$E_{intra} = (Q_{enlazado}^{L-L} - Q_{no\ enlazado}^{L-L}) + (Q_{enlazado}^{R-R} - Q_{no\ enlazado}^{R-R}) \quad (1)$$

La energía intermolecular tiene en cuenta los estados de enlace y no enlace respecto al complejo completo:

$$E_{inter} = (Q_{enlazado}^{R-L} - Q_{no\ enlazado}^{R-L}) \quad (2)$$

donde Q se define de la siguiente manera:

$$Q = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{enlaces\ H} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) \quad (3)$$

Esta ecuación integra los términos de energía asociados a las fuerzas de dispersión/repulsión (vdw), enlaces de hidrógeno (enlaces H), enlaces electrostáticos (elec), enlaces de solvatación (sol). Los términos W_{vdw} , $W_{enlaces\ H}$, W_{elec} y W_{sol} corresponden a constantes asociadas a las fuerzas de Van der Waals, enlaces de hidrógenos, interacciones electrostáticas y de solvatación, respectivamente. La distancia interatómica, representada mediante r_{ij} y los parámetros de Lennard-Jones tomados de la fuerza A_{ij} , B_{ij} , C_{ij} y D_{ij} se utilizan para las energías potenciales entre dos átomos y $E(\theta)$ define el ángulo de direccionalidad (60).

2.6.1.2 Acoplamiento molecular de complejos péptido-proteína

La flexibilidad que caracteriza a un péptido representa el mayor reto a la hora de realizar un acoplamiento molecular que involucre a este tipo de moléculas. Incluso en el caso donde el sitio activo o de unión en el receptor es bien conocido e incluso si se cuenta con información sobre la localización y conformación aproximada del péptido, la identificación de la conformación correcta involucra un espacio de búsqueda enorme. Una manera de abordar este problema es segmentar el acoplamiento del péptido en múltiples “módulos” independientes, con el fin de enfocarse en las conformaciones más relevantes y, de esta manera, reducir el espacio de búsqueda. De acuerdo a lo que se ha podido aprender sobre las características conformacionales de péptidos cortos, se suele asumir que la conformación de un péptido fluctúa en equilibrio entre una cantidad más o menos restringida de conformaciones “preferidas” y una unión resulta en un cambio en dicho equilibrio hacia una conformación de enlace preexistente. Por lo tanto para el caso del péptido, la proyección de un conjunto conformacional puede reducir significativamente los grados de libertad internos que deben ser explorados para el péptido durante el modelamiento del acople molecular. Por consiguiente, el espacio conformacional del esqueleto peptídico (conocido usualmente como *backbone*) puede reducirse en gran medida mediante la consideración de un número limitado y específico de conformaciones tales como la de la prolipolina extendida o en hélice, o fragmentos extraídos de estructuras de proteínas bien conocidas. Para el caso del receptor, la búsqueda puede restringirse a un conjunto limitado de sitios activos prometedores previamente identificados (61).

De acuerdo a las simplificaciones previamente mencionadas, modelos iniciales de un complejo péptido-receptor se pueden generar de manera eficiente. No obstante, estos modelos iniciales requieren de un mayor refinamiento para aumentar la calidad de los resultados. Este último proceso de acople y optimización puede conducir en cambios conformacionales significativos del péptido, sobre todo si en el modelamiento se considera el fenómeno de pliegue y enlace del péptido de manera simultánea. De esta manera, un protocolo de acople molecular debe ser capaz de muestrear los cambios conformacionales requeridos de manera tal que se evalúen conformaciones

cercanas a la nativa y que estas sean refinadas adecuadamente hacia modelos que minimicen la energía del complejo (61).

2.6.2 Rosetta FlexPepDock

Rosetta FlexPepDock (62, 63) es un protocolo computacional que permite modelar el acoplamiento entre una proteína y un péptido. Esta herramienta fue una de las primeras en abordar el reto del vasto espacio conformacional que implican los péptidos, pues logró modelar con precisión péptidos de una longitud superior a 5 aminoácidos acoplados a un sitio activo de un receptor, utilizando una combinación de acople ligando-receptor (optimización de cuerpo rígido) y pliegue del ligando (optimización del *backbone* del péptido).

Este protocolo hace parte de la *suite* de Rosetta (64–66), el cual está basado en el método Monte-Carlo para el muestreo del espacio conformacional relevante, utilizando una función de energía calibrada especialmente para diversas aplicaciones de modelado y diseño molecular.

Modelos precisos de complejos péptido-proteína pueden ser utilizados como puntos de partida para subsiguientes caracterizaciones detalladas de las interacciones en el complejo. La comprensión de la especificidad en la unión de un péptido a cierto receptor puede ayudar a predecir la compatibilidad de unión a un sitio activo dado.

2.6.2.1 Descripción del protocolo Rosetta FlexPepDock

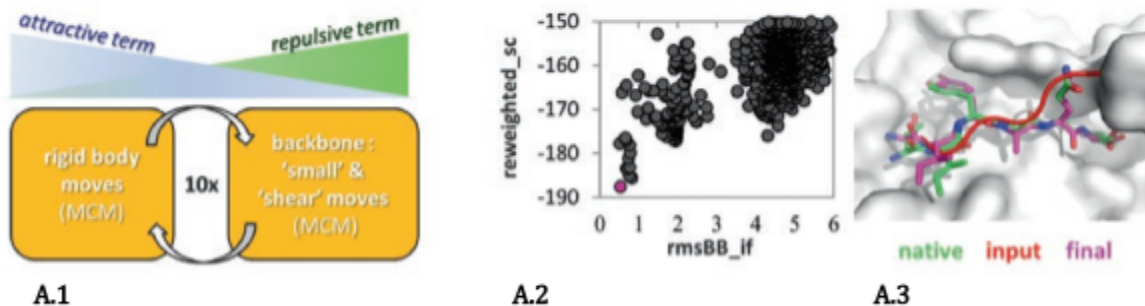
El protocolo FlexPepDock cuenta con dos variantes (ver **Figura 2-7**):

1. ***FlexPepDock Refinement***: Este protocolo está diseñado para realizar refinamientos de complejos a través de múltiples ciclos iterativos de optimizaciones (a través del método Monte-Carlo) de las componentes de cuerpo rígido (traslaciones y rotaciones de cuerpo rígido) y del *backbone* (ángulos diedros) del péptido. A lo largo de dichos ciclos se van modificando los términos correspondientes a las fuerzas repulsivas (se reducen) y atractivas (se incrementan) de Van der Waals en la función de energía, los cuales van cambiando gradualmente hasta llegar a los valores originales. Esto permite realizar un barrido eficiente de las configuraciones de mayor energía, para que luego se exploren con mayor precisión las configuraciones de menor energía. Todas las conformaciones derivadas de los múltiples ciclos de optimización son evaluadas a través de la función de energía y se retiene la configuración de menor energía. Todas las cadenas laterales en la interfaz del complejo son continuamente reajustadas durante el proceso de refinamiento, esto

permitiendo sustituciones entre el conjunto de posibles rotámeros asociados a cada residuo.

2. **FlexPepDock *ab-initio*:** Este protocolo permite explorar un espacio mucho más amplio de conformaciones del péptido. Esto es particularmente útil cuando no se tiene información previa de la conformación del péptido. El protocolo inicia con una serie de ciclos iterativos de optimización de las componentes de cuerpo rígido y del *backbone* del péptido, en donde la estructura es reemplazada por un modelo de “grano grueso” en donde las cadenas laterales son sustituidas por centroides. Estos ciclos inician a alta temperatura (en el sentido del método Monte-Carlo), la cual es reducida gradualmente hasta llegar a la temperatura normal. Luego cada modelo resultante del paso anterior es refinado en el marco de un modelo donde se consideren todos los átomos a través del protocolo *FlexPepDock Refinement*. Todas las conformaciones derivadas del protocolo previamente descrito son evaluadas a través de una función de energía y son agrupadas con el fin de seleccionar los modelos más representativos.

A. Protocolo *FlexPepDock Refinement*: Refinamiento de modelos a nivel atómico



B. Protocolo *FlexPepDock ab-initio*: Pliegue y acople modelo centroides + refinamiento a nivel atómico = modelo final

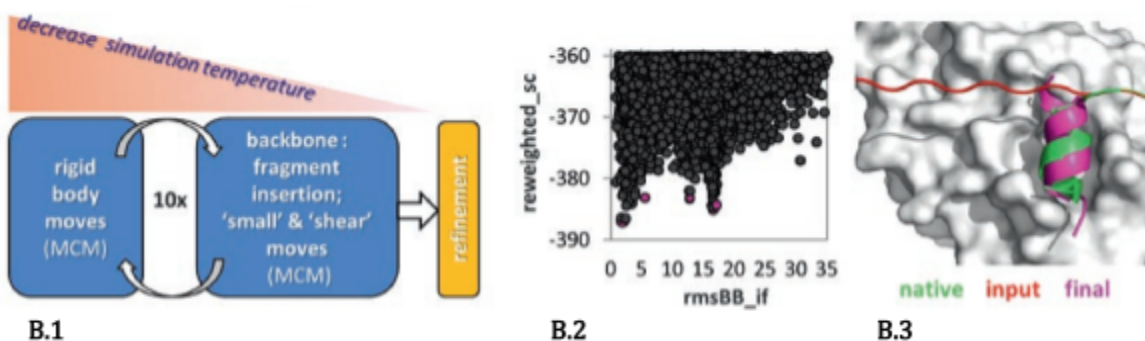


Figura 2-7: Descripción general de los protocolos Rosetta FlexPepDock. **(A.1)** Esquema descriptivo del protocolo de refinamiento *Rosetta FlexPepDock Refinement*. La estructura inicial se refina a través de ciclos iterativos de optimización (minimización) de los componentes de cuerpo rígido del complejo y *backbone* del péptido. A lo largo de los ciclos de optimización, se varían los valores de las fuerzas repulsivas y atractivas de Van der Waals. **(A.2)** Gráfica de rmsBB_if (desviación estándar de los átomos que conforman el *backbone* del péptido de residuos en la interfaz de unión) vs reweighted_sc (función de puntuación de la energía ponderada) de los resultados de 1000 simulaciones donde el punto magenta corresponde al modelo de menor energía. **(A.3)** Ejemplo de péptido unido al dominio SH2 de la proteína XLP SAP (67). El modelo inicial se muestra en color rojo, el modelo final refinado en magenta y la estructura resuelta experimentalmente en verde. **(B.1)** Esquema descriptivo del protocolo *Rosetta FlexPepDock ab-initio*. La estructura inicial (con el péptido en alguna conformación arbitraria) es refinada a través de múltiples ciclos iterativos de optimización de las componentes de cuerpo rígido y del *backbone* del péptido, utilizando un modelo donde se reemplazan los residuos por centroides. Estos ciclos inician con una temperatura de simulación alta, la cual se va reduciendo paulatinamente con el pasar de los ciclos. Cada modelo resultante es refinado nuevamente a través del protocolo *Rosetta FlexPepDock Refinement* utilizando un modelo completamente atómico. **(B.2)** Gráfica de los resultados de 50.000 simulaciones. Los puntos en magenta representan los modelos más representativos y de menor energía. **(B.3)** Ejemplo de un péptido derivado del receptor nuclear coactivador 1 unido al receptor de mineralocorticoides. El péptido nativo se presenta en color verde, el inicial en rojo y el final (optimizado) en magenta (68). Abreviaciones: MCM = Monte-Carlo Minimization. Imágenes tomadas de (61).

3. Materiales y métodos

Los datos utilizados para la implementación y estandarización de la estrategia de identificación y priorización de neo-antígenos candidatos a vacuna personalizada contra el cáncer, consisten en un conjunto de secuenciaciones obtenidas de una paciente con PEComa.

3.1 Resumen de historia clínica de la paciente

Paciente de 63 años de edad con diagnóstico de PECOMA desde hace 2 años, con múltiples lesiones en cerebro, hígado, riñón, tratadas quirúrgicamente y con posterior medicación con everolimus, quien no ha mostrado respuesta adecuada al medicamento. En estudio por PET CT se evidenciaron múltiples lesiones activas en columna (C1, T12, L1 y L3), base pulmonar derecha (9mm), hígado (37mm), riñón izquierdo asociado a masa necrótica (90x85x105mm) y lesión iliaca izquierda. De acuerdo a estos resultados, se concluye una enfermedad progresiva (comparado con estudio previo) por lesiones nuevas que son metabólicamente activas.

3.2 Datos ómicos

A partir de una muestra de una lesión tumoral de columna obtenida quirúrgicamente (tumor) y de una muestra de sangre (células sanas), se procedió a secuenciar el exoma y transcriptoma del tejido tumoral y el exoma de células sanas.

En la **Tabla 3-1** se resumen las principales características del conjunto de datos de secuenciaciones utilizadas en la presente investigación. Dichas secuenciaciones fueron realizadas por la empresa Novogene.

Tipo secuenciación	Exoma células normales	Exoma células tumorales	Transcriptoma células tumorales
Origen células secuenciadas	Sangre periférica	columna vertebral	columna vertebral
Profundidad	150x	300x	100M lecturas
% lecturas con Q_{Phred}^{30}	92,34	92.13	92.75
% lecturas con Q_{Phred}^{20}	96.89	96.77	97.07
% GC	50.95	50.59	53.36
Tamaño lectura	150 pb	150 pb	150 pb
Tecnología secuenciación	Illumina SBS technology, Pair End	Illumina SBS technology, Pair End	Illumina SBS technology, Pair End

Tabla 3-1: Características de las secuenciaciones utilizadas para implementar y estandarizar la estrategia in-silico de identificación y priorización de epítopes potencialmente inmunogénicas.

Basándose en la metodología planteada en (1), en la **Figura 3-1** se resume la metodología implementada en el presente trabajo, la cual se compone de 8 etapas:

1. Control de calidad. Se utilizó FastQC (69). Para procesamientos intermedios (adecuaciones de formato) se utilizó samtools (70), picard (71) y bcftools (72, 73).
2. Alineamiento por referencia de los exomas tumoral y normal y del transcriptoma: Se utilizó la herramienta BWA (74) para los exomas y tophat2 (75) para el transcriptoma. Se utilizó como genoma de referencia la versión GRCh38/hg38.
3. Identificación de variantes: Se comparan los exomas normal y tumoral con el fin de identificar las mutaciones somáticas. Se utilizó la herramienta Mutect2 de GATK (76).
4. Anotación de variantes: se realiza una anotación de las mutaciones identificadas, es decir, se realiza una búsqueda en múltiples bases de datos para obtener información sobre cada mutación (nombre del gen donde se ubica la mutación, transcritos asociados a dicho gen, tipo de mutación, etc.). Se utilizó la herramienta VEP (46).

5. Tipificación del haplotipo HLA (12): Consiste en realizar un alineamiento respecto a una base de datos que contiene las secuencias de los alelos HLA. Se utilizó la herramienta optiType (47) para el HLA clase I y seq2HLA (48) para el HLA clase II.
6. Cuantificación de los niveles de expresión de los transcritos codificados en el transcriptoma tumoral: Se cuantifican las lecturas asociadas a cada transcrito, esta cuantificación se da comúnmente en unidades conocidas como FPKM (*fragments per kilobase of transcript per megabase of sequenced reads*). Se utilizó la herramienta cufflinks (77).
7. Predicción de epítopes: se estima el procesamiento del péptido candidato a vacuna por parte del proteasoma y su presentación según su afinidad y estabilidad en el complejo péptido-MHC. Se utilizó la herramienta pvacseq (1).
8. Priorización de epítopes: Se realiza una selección de las mejores epítopes candidatas a vacuna para el alelo HLA-A*24:02 (para este alelo existen estructuras resueltas por difracción de rayos X en la base de datos PDB (55)), a los cuales se les realiza *docking* molecular para obtener información que permita priorizar los mejores candidatos a vacuna. Se utilizó la herramienta Rosetta, en particular el protocolo FlexPepDock Ab-initio (62). En la **Figura 3-2** se resume el protocolo implementado para el docking molecular.



Figura 3-1: Flujo de trabajo de la identificación y priorización de neoantígenos.

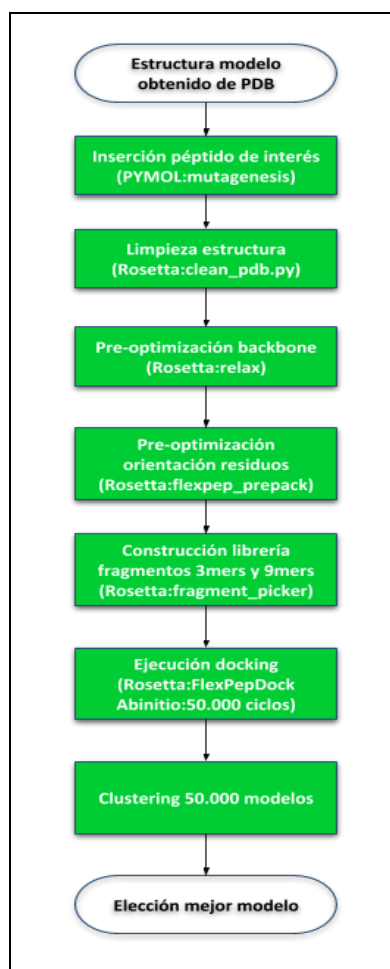


Figura 3-2: Etapas del protocolo implementado para la ejecución del *docking* molecular.

4. Resultados y Discusión

Una vez implementadas y ejecutadas las diferentes etapas de la metodología propuesta, se realizaron los respectivos análisis los cuales permitieron identificar y priorizar epítopes candidatas a vacuna personalizada. A continuación se presentan los diferentes resultados obtenidos en cada una de dichas etapas:

Etapas 1: Control de calidad

El análisis con la herramienta FastQC se hizo sobre los dos conjuntos de lecturas para cada secuenciación (un conjunto denominado *Forward* y el otro *Reverse* debido a la secuenciación tipo Pair-End).

Para el caso del exoma normal, en lo que respecta a las lecturas *Forward* se obtuvieron resultados muy satisfactorios pues el valor promedio del indicador de calidad Q_{phred} por posición en la lectura se ubicó alrededor de 39 (**Figura 4-1**). Para el caso de las lecturas *Reverse* se obtuvieron resultados satisfactorios aunque inferiores a los correspondientes para las lecturas *Forward*, pues el valor promedio del indicador de calidad Q_{phred} por posición en la lectura se ubicó alrededor de 34, observándose un ligero deterioro en los valores de las posiciones finales de la lectura (**Figura 4-2**). Estos resultados combinados a la importante profundidad con la que se secuenció este exoma (150X) otorgarán una alta confianza a los resultados de identificación de variantes.

Para el caso del exoma tumoral, en lo que respecta a las lecturas *Forward* se obtuvieron resultados muy satisfactorios pues el valor promedio del indicador de calidad Q_{phred} por posición en la lectura se ubicó alrededor de 38 (**Figura 4-3**). Para el caso de las lecturas *Reverse* se obtuvieron resultados satisfactorios aunque inferiores a los correspondientes para las lecturas *Forward*, pues el valor promedio del indicador de calidad Q_{phred} por posición en la lectura se ubicó alrededor de 33, observándose una mayor variabilidad en los valores y un ligero deterioro en las posiciones finales de la lectura (**Figura 4-4**). Estos resultados combinados a la importante profundidad con la que se secuenció este exoma (300X ya que se asume una “contaminación” del 50% de células normales en la muestra) otorgarán una alta confianza a los resultados de identificación de variantes.

Para el caso del transcriptoma tumoral, en lo que respecta a las lecturas *Forward* se obtuvieron resultados muy satisfactorios pues el valor promedio del indicador de calidad Q_{Phred} por posición en la lectura se ubicó alrededor de 38 (**Figura 4-5**). Para el caso de las lecturas *Reverse* se obtuvieron resultados satisfactorios aunque inferiores a los correspondientes para las lecturas *Forward*, pues el valor promedio del indicador de calidad Q_{Phred} por posición en la lectura se ubicó alrededor de 33, observándose una mayor variabilidad en los valores y un ligero deterioro en las posiciones finales de la lectura (**Figura 4-6**). Estos resultados combinados a la importante cantidad de lecturas con la que se secuenció este transcriptoma (100M lecturas) otorgarán una alta confianza a los resultados de cuantificación de los niveles de expresión de las variantes identificadas.

En el **Anexo 1** se presentan los resultados completos del análisis FastQC.

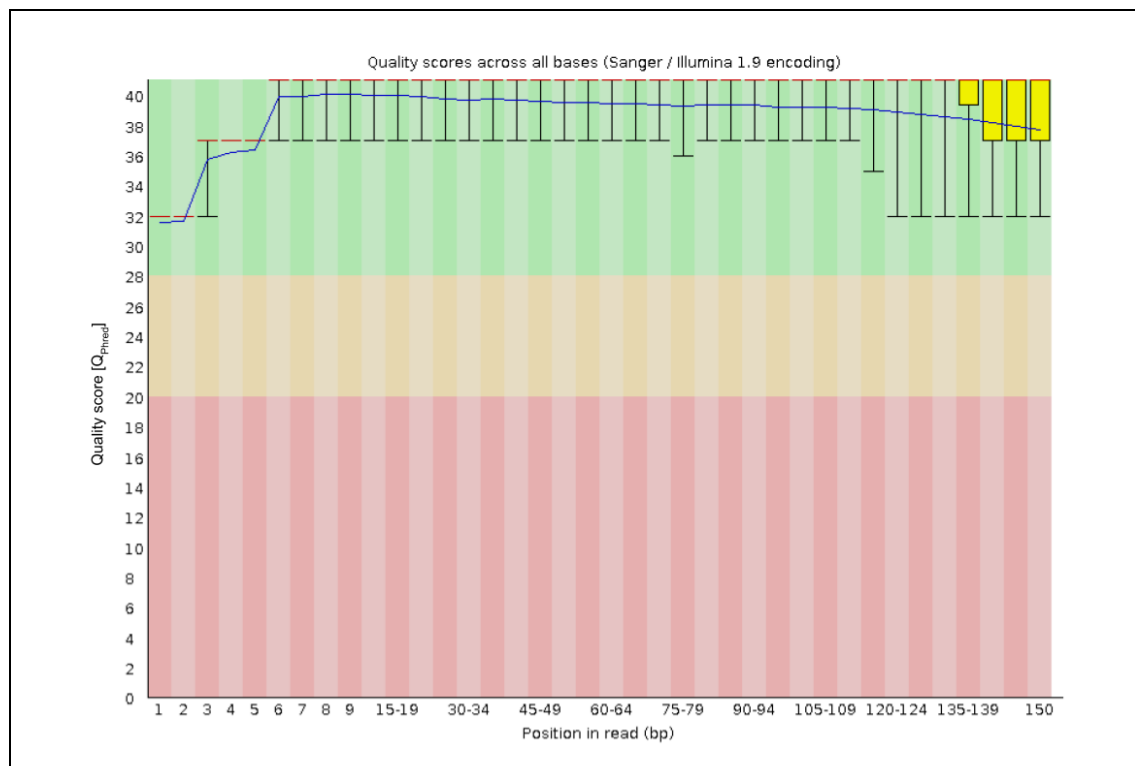


Figura 4-1: Puntajes promedio de calidad [Q_{Phred}] por posición en la lectura para el exoma normal (lecturas *Forward*).

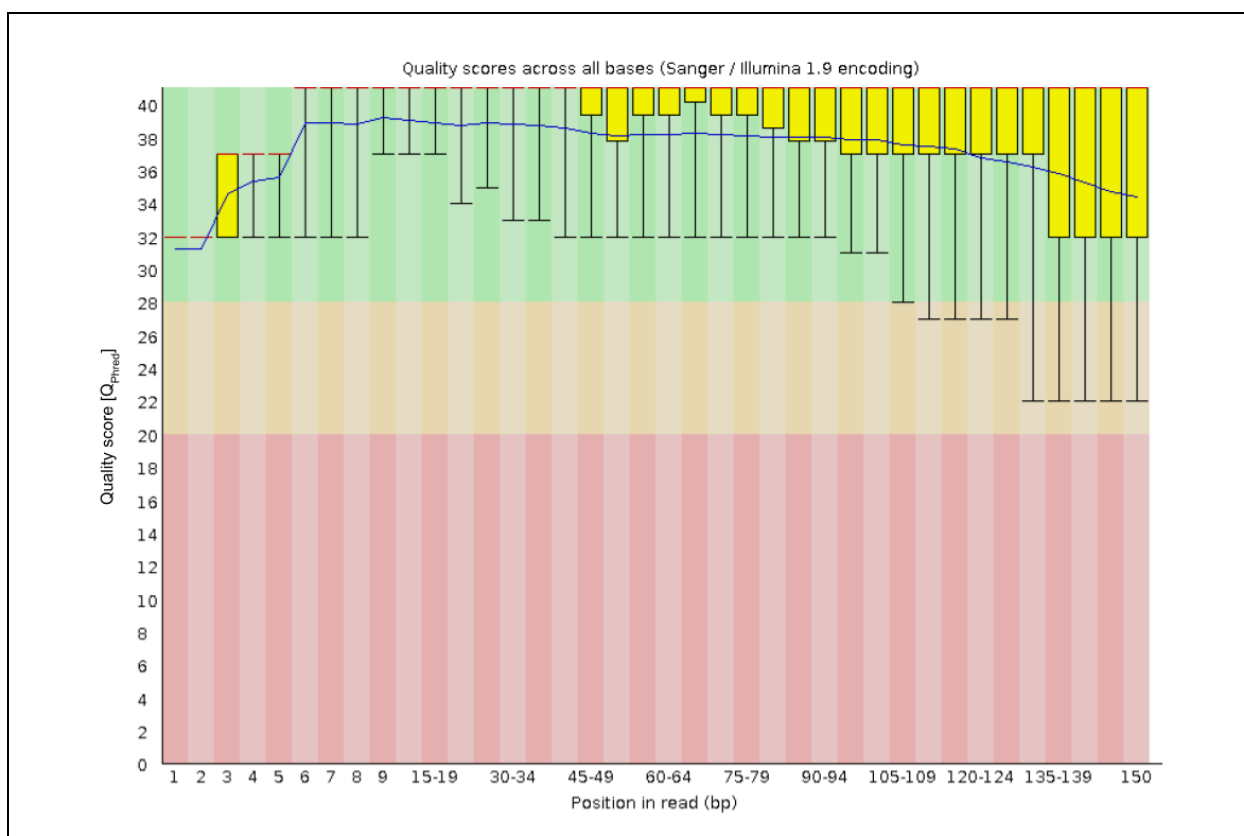


Figura 4-2: Puntajes promedio de calidad [Q_{phred}] por posición en la lectura para el exoma normal (lecturas *Reverse*).

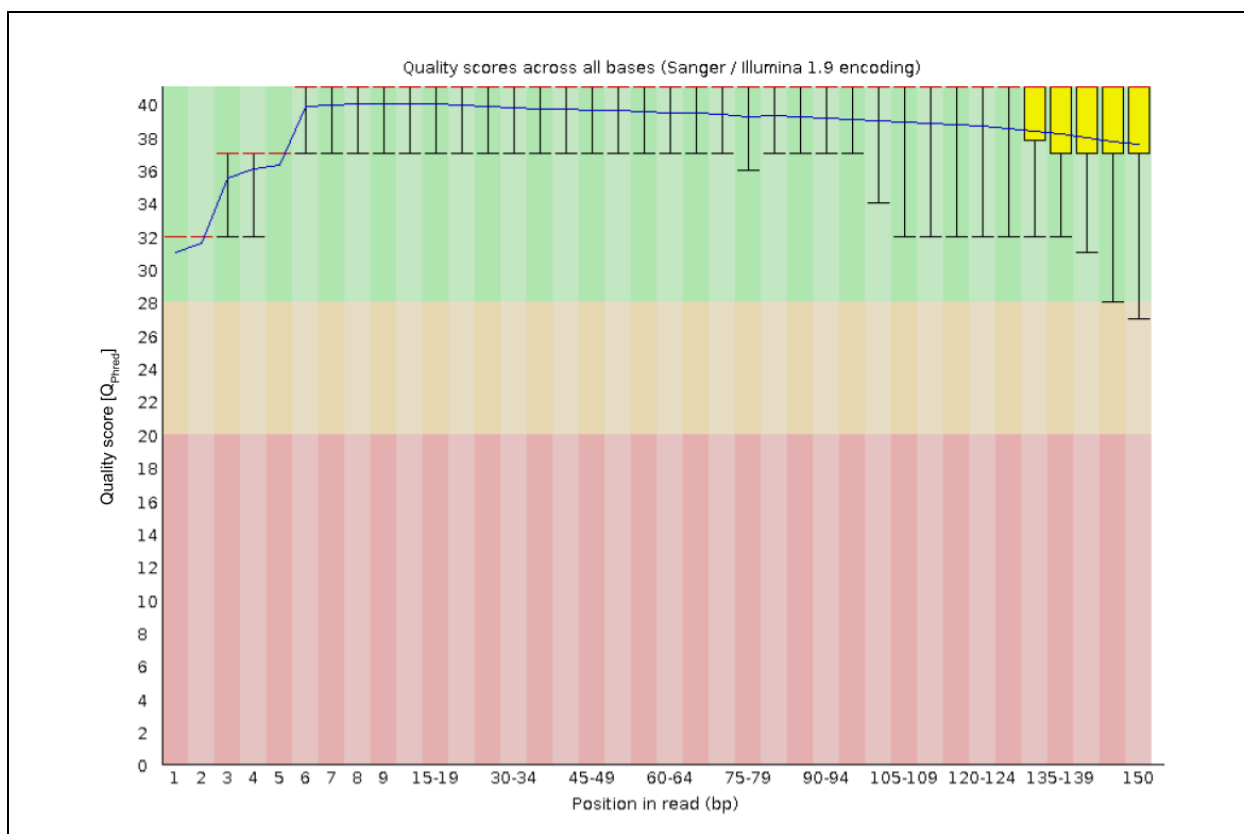


Figura 4-3: Puntajes promedio de calidad [Q_{phred}] por posición en la lectura para el exoma tumoral (lecturas *Forward*).

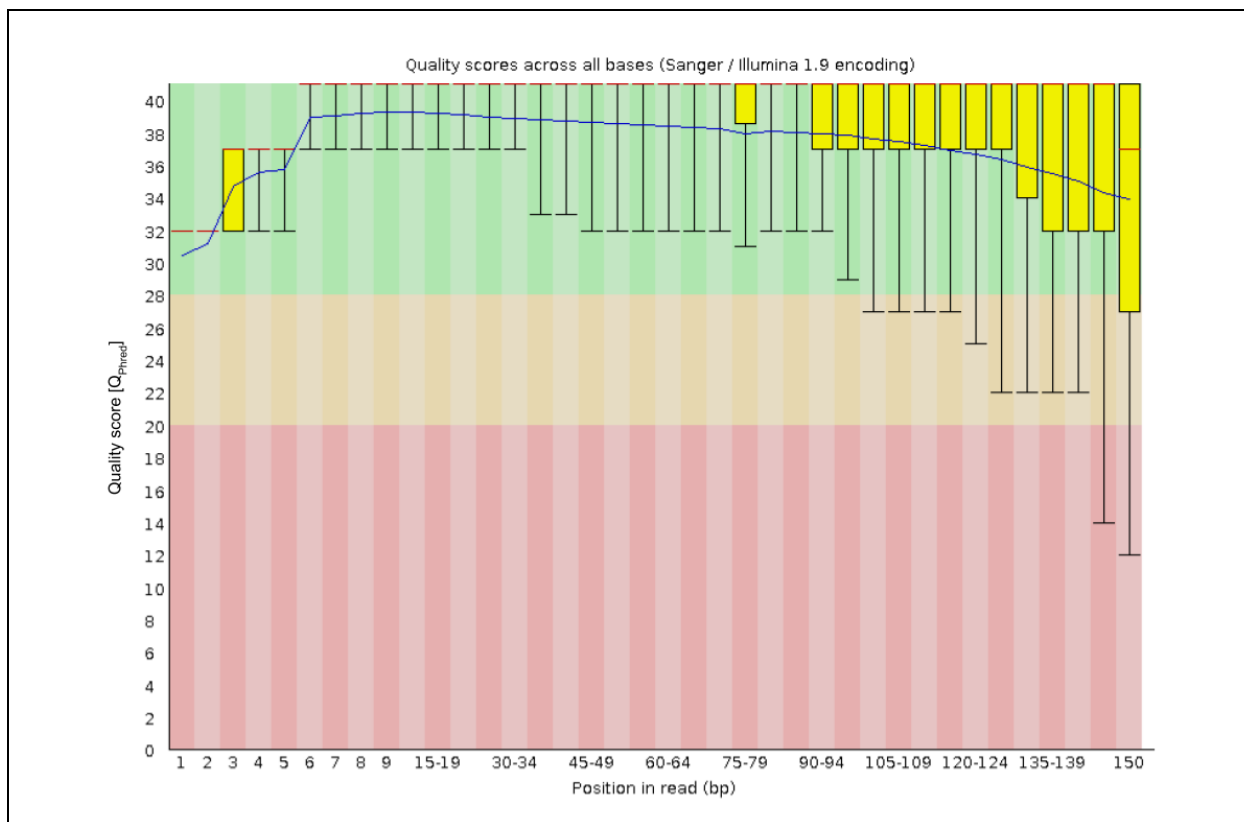


Figura 4-4: Puntajes promedio de calidad [Q_{phred}] por posición en la lectura para el exoma tumoral (lecturas *Reverse*).

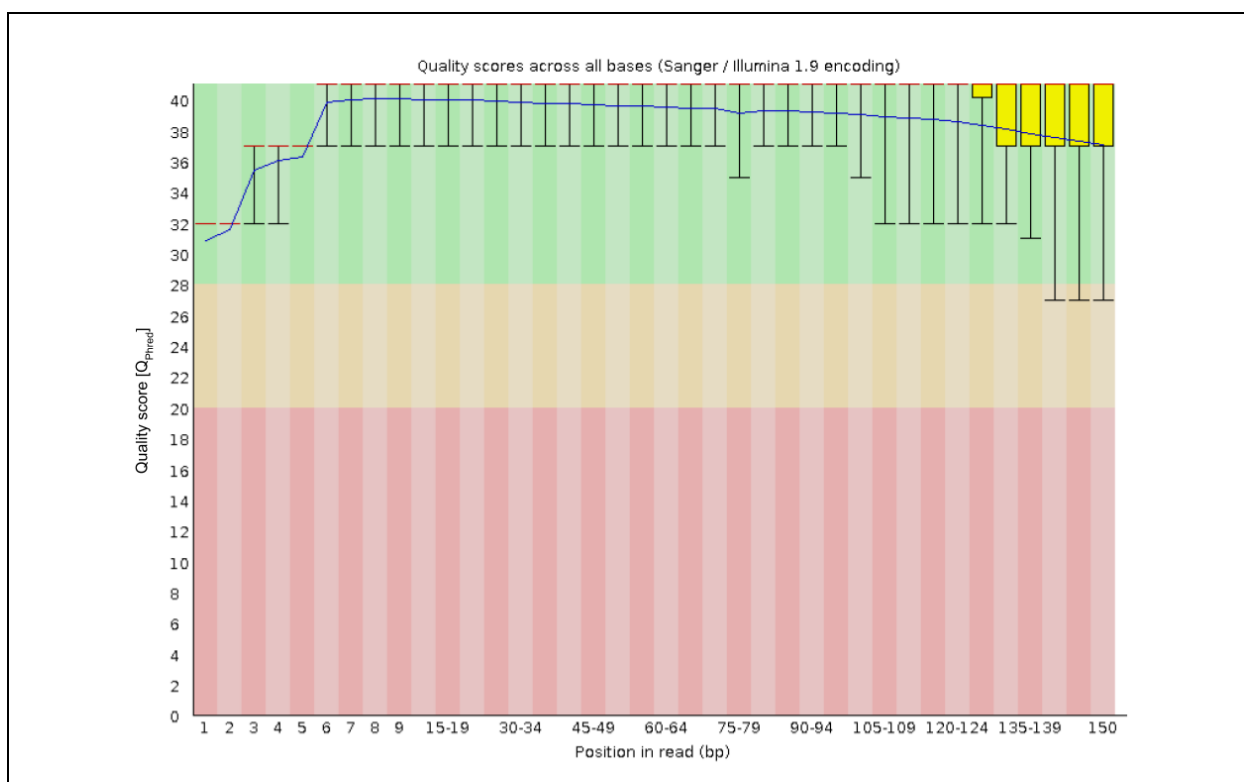


Figura 4-5: Puntajes promedio de calidad [Q_{phred}] por posición en la lectura para el transcriptoma tumoral (lecturas *Forward*).

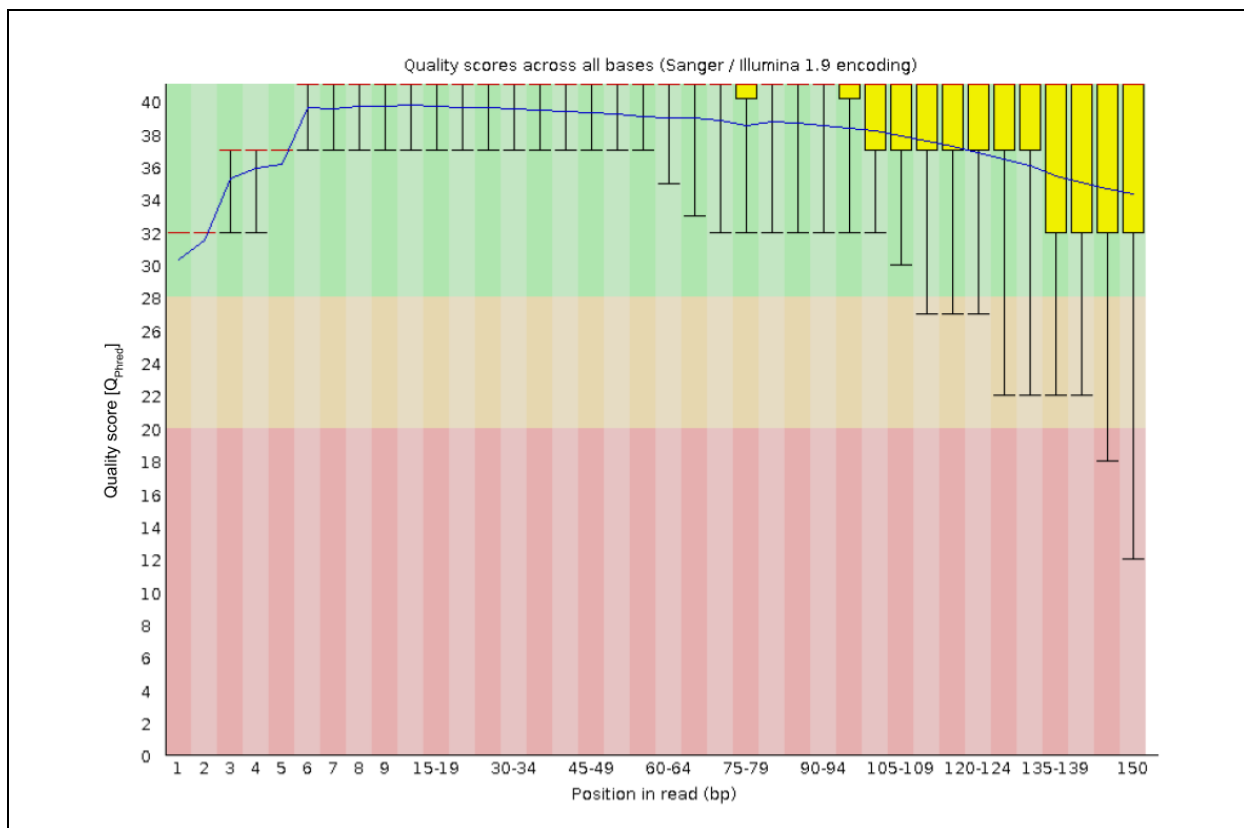


Figura 4-6: Puntajes promedio de calidad [Q_{phred}] por posición en la lectura para el transcriptoma tumoral (lecturas *Reverse*).

Etapas 2: Alineamientos.

Exoma completo tejido normal: De las 70'139.549 de lecturas analizadas, 227.023 de lecturas no se pudieron mapear (0,32%).

Exoma completo tejido tumoral: De las 120'041.101 de lecturas analizadas, 305.191 de lecturas no se pudieron mapear (0,25%).

Transcriptoma completo tejido tumoral: De las 144'123.061 pares de lecturas analizadas, en promedio el 82,5% se pudo mapear.

Etapas 3: Identificación de variantes.

Se identificaron 23.314 variantes las cuales involucran 14.264 genes.

Etapas 4: Anotación de variantes.

En la **Figura 4-7** se presenta la distribución de las variantes de acuerdo con las siguientes categorías: Inserción, deleción y SNV (Single Nucleotide Variant), en donde se puede resaltar la

significativa cantidad de indels. En la **Figura 4-8** se muestra la distribución de las variantes de acuerdo con el tipo de consecuencia en regiones codificantes, en donde se resalta la alta proporción de variantes de cambio de un solo aminoácido (*missense variants*) y la importante proporción de variantes sinónimas (*synonymous variants*). En la **Figura 4-9** se muestra la distribución de variantes por cromosoma, en donde se identifica una importante cantidad de variantes en el cromosoma 1. En la **Figura 4-10** se muestra la distribución de las variantes de acuerdo con la posición en la que se ubican en la proteína traducida (indicada en rangos porcentuales).

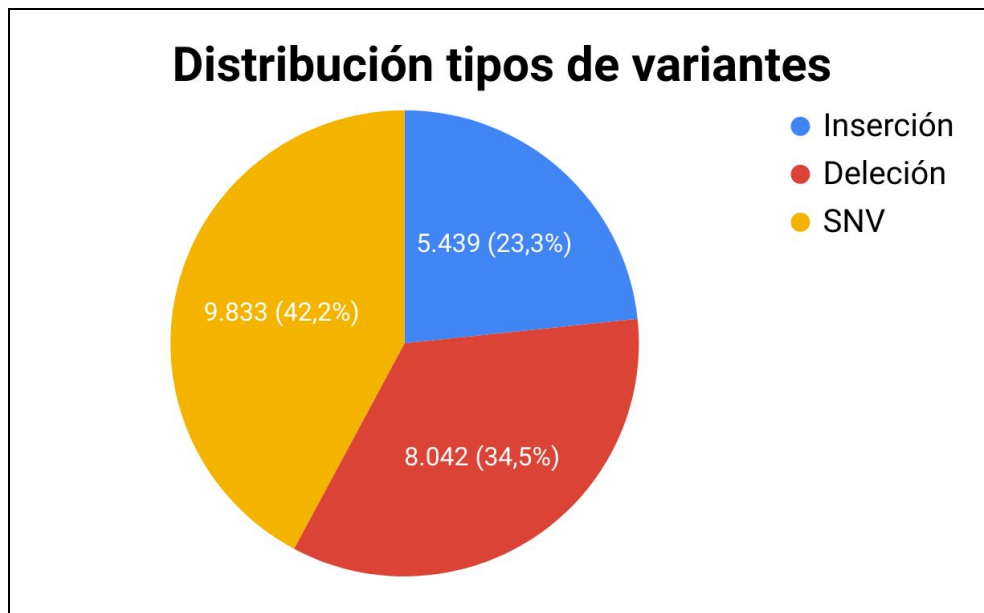


Figura 4-7: Distribución de las variantes de acuerdo con su tipo.

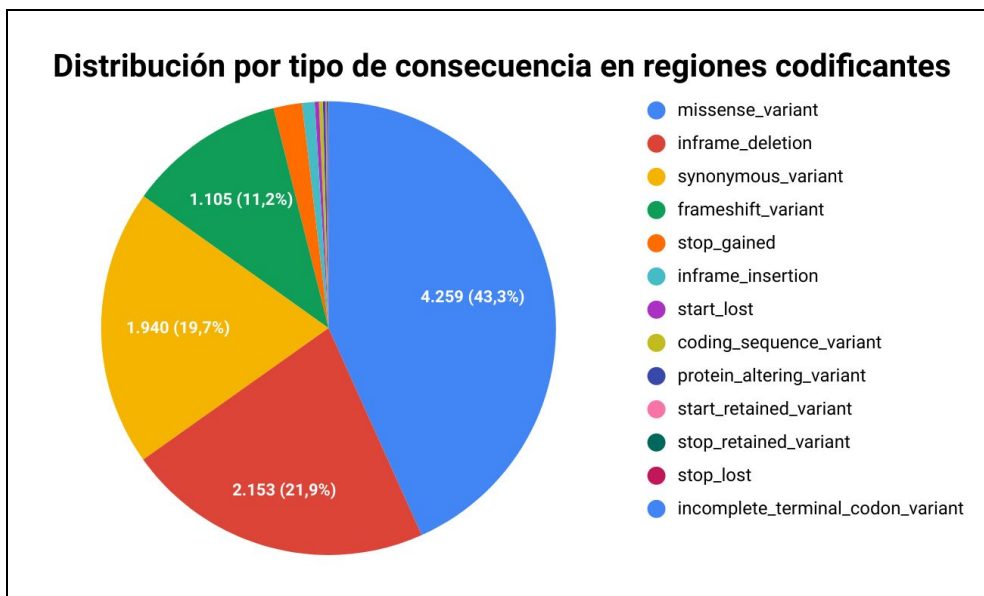


Figura 4-8: Distribución de las variantes de acuerdo al tipo de consecuencia en regiones codificantes.

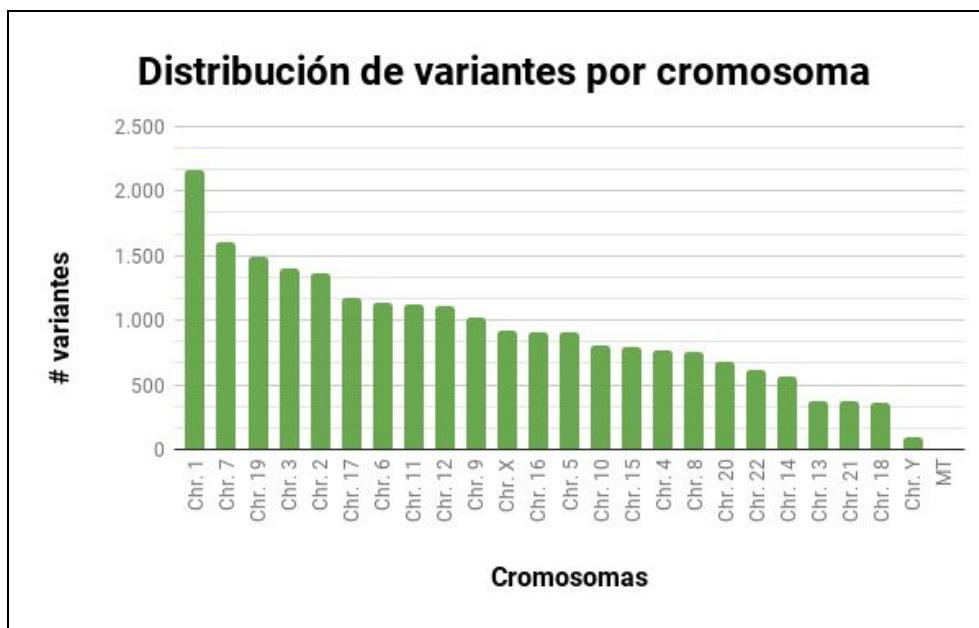


Figura 4-9: Distribución de variantes por cromosoma.

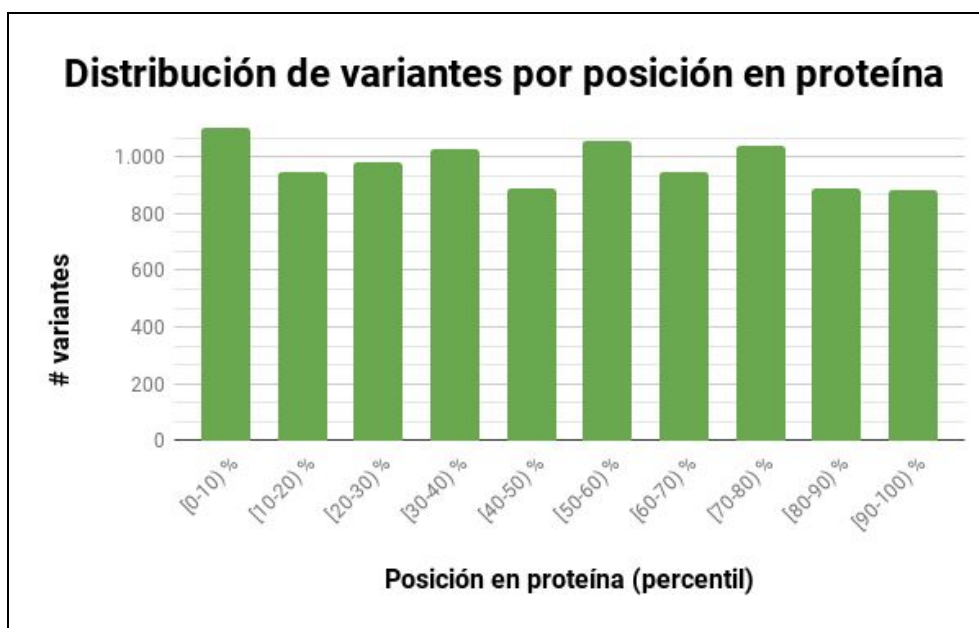


Figura 4-10: Distribución de variantes por posición en proteína.

Etapas 5: Tipificación del haplotipo HLA.

En la **Tabla 4-1** se resume el haplotipo HLA identificado.

Clase I	Clase II
A*24:02	DQA1*06:01
A*31:01	DQA1*06:01
B*51:01	DQB1*03:09
B*51:01	DQB1*03:09
C*14:02	DRA*01:01
C*14:02	DRA*01:01
	DRB1*08:02
	DRB1*12:02
	DPA1*02:02
	DPA1*02:02
	DPB1*100:01
	DPB1*05:01

Tabla 4-1: Haplotipo HLA de la paciente

En el **Anexo 2** se presenta información suplementaria sobre los resultados de tipificación HLA.

Etapas 6: Cuantificación de los niveles de expresión.

Se identificaron 22.688 genes y 123.563 transcritos expresándose en el transcriptoma del tejido tumoral.

Etapas 7: Predicción de epítopes.

Se identificaron 12 péptidos candidatos a vacuna, 6 de los cuales son péptidos largos (entre 18 y 21 aminoácidos) que contienen tanto una epítope para clase I como para clase II. Los otros 6 péptidos son péptidos cortos (entre 9 y 10 aminoácidos) dirigidos a clase I. En la **Tabla 4-2** se enlistan los péptidos identificados.

Secuencia larga MT	Secuencia corta MT	Secuencia corta WT
RPLWAW FQ RARVQLRRDLR	FQ RARVQLR	FQ RARVSSS
KPYIYA GYR PLYQ TW RFYFRT	GYRPLYQ TW	GYRPLHQ TW
YAGYR PLYQ T WRFYF RTL	LYQ TW RFYF	LHQ TW RFYF
LAGTLG RMLAV T FLL QA	RMLAVT FLL	SMLAVT FLL
NA	ISKKKHYNR	ISKKKALQQ
NA	RQQQQHLHR	QQQQQHLHR
NA	RQQQQQILR	QQQQQQILR
NA	RMRMSSTPPR	NA
YAGYR PLYQ T WRFYF RTL FQ	PLYQ TW RFYF	PLHQ TW RFYF
CKDLI KF MLRNER QF KEEKL	KFMLRNER QF	KSMLRNER QF
NA	LWARRCPWTM	NA
NA	HFKA P KISMR	HFKA P KISMP

Tabla 4-2: Lista de péptidos candidatos a vacuna (MT: Mutant, WT: WildType).

En el **Anexo 3** se presentan los detalles de la predicción de los péptidos candidatos a vacuna.

Etapas 8: *Docking* molecular.

De acuerdo a los resultados resumidos en la **Tabla 4-2**, para el *docking* molecular se seleccionan los péptidos que son presentados por la molécula HLA-A*24:02 debido a que de esta se tienen estructuras resueltas por difracción de rayos X. Para los péptidos de 9 aminoácidos de longitud se utilizó como plantilla la estructura 2BCK (resolución: 2.8 Å)(55), mientras que para los péptidos de 10 aminoácidos se utilizó la estructura 3NFN (resolución: 2.392 Å)(55). Igualmente se opta por simular únicamente los péptidos cortos dirigidos a clase I debido a las limitaciones de plantillas para péptidos largos y la limitante en capacidad de cálculo disponible (cada caso de simulación requería entre 8 y 10 días utilizando el 100% del servidor). Por lo anterior, se escogieron 5 péptidos a simular, de los cuales también se simuló su contraparte nativa. En la **Tabla 4-3** se resumen los péptidos simulados.

Secuencia larga MT	Secuencia corta MT	Secuencia corta WT	Longitud péptido corto
YAGYR LYQTWRFYF RTRL	LYQTWRFYF	LHQTWRFYF	9
LAGTLG RMLAVTFLL QA	RMLAVTFLL	SMLAVTFLL	9
KPYIYA GYRPLYQ TWRFYFRT	GYRPLYQTW	GYRPLHQTW	9
YAGYR PLYQTWRFYF RTRLFQ	PLYQTWRFYF	PLHQTWRFYF	10
CKDLI KFMLRNERQF KEEKL	KFMLRNERQF	KSMRLNERQF	10

Tabla 4-3: Lista de péptidos simulados (MT: Mutant, WT: WildType). En fondo azul los péptidos a los que se les realizó *docking* molecular con la molécula HLA-A*24:02.

Con el fin de verificar que el *backbone* del receptor (la proteína HLA) no haya sufrido cambios significativos durante el *docking*, se calculó el RMSD (*Root-Mean-Square Deviation*) entre el *backbone* del receptor obtenido por el *docking* molecular y el *backbone* del receptor de la plantilla correspondiente. En la **Tabla 4-4** se resumen los resultados obtenidos, los cuales confirman que no se presentaron cambios significativos en el *backbone* del receptor durante el *docking* molecular.

Secuencia corta MT	RMSD [Å]	Secuencia corta WT	RMSD [Å]
LYQTWRFYF	1,150	LHQTWRFYF	1,142
RMLAVTFLL	1,257	SMLAVTFLL	1,221
GYRPLYQTW	1,129	GYRPLHQTW	1,207
PLYQTWRFYF	1,127	PLHQTWRFYF	1,216
KFMLRNERQF	1,181	KSMRLNERQF	1,189

Tabla 4-4: Cálculo del RMSD entre el backbone del receptor resultante del docking y el backbone del receptor de la plantilla respectiva para todos los complejos simulados.

De los resultados del *docking* molecular se pueden extraer 2 características que pueden proporcionar información relevante sobre la interacción péptido proteína y por ende contribuir a la priorización de los péptidos candidatos a vacuna. La primera son los puentes de hidrógeno y la segunda es el valor de la función de puntaje obtenida durante el *docking* molecular.

En la **Tabla 4-5** se resumen los resultados de la identificación de puentes de hidrógenos en la interface péptido-HLA para todos los complejos simulados utilizando el software Chimera (59). En

el **Anexo 4** se presentan figuras que permiten visualizar los puentes de hidrógeno identificados, así como otras figuras obtenidas del *docking* molecular.

Secuencia corta MT	Puentes H	Secuencia corta WT	Puentes H
LYQTWRFYF	10	LHQTWRFYF	10
RMLAVTFLL	10	SMLAVTFLL	11
GYRPLYQTW	9	GYRPLHQTW	5
PLYQTWRFYF	14	PLHQTWRFYF	14
KFMLRNERQF	10	KSMLRNERQF	16

Tabla 4-5: Identificación de puentes de hidrógeno en la interface péptido-HLA.

En la **Tabla 4-6** se presentan los valores obtenidos por la función de puntuación utilizada por Rosetta para cada complejo simulado. Vale la pena aclarar que las unidades de la función de puntuación de Rosetta no son unidades de energía, ya que Rosetta, además de calcular los términos de energía derivados de las interacciones interatómicas, añade algunos términos para darle mayor peso a las interacciones que se presentan en la interface péptido-proteína con el fin de mejor dirigir las optimizaciones de la simulación. Por lo cual, formalmente se debe hablar de unidades arbitrarias aún cuando estos valores sean directamente proporcionales a la energía.

Secuencia corta MT	Puntaje [REU*]	Secuencia corta WT	Puntaje [REU*]
LYQTWRFYF	-661.156	LHQTWRFYF	-643.907
RMLAVTFLL	-663.843	SMLAVTFLL	-672.304
GYRPLYQTW	-677.147	GYRPLHQTW	-659.998
PLYQTWRFYF	-737.283	PLHQTWRFYF	-734.283
KFMLRNERQF	-725.444	KSMLRNERQF	-723.828

Tabla 4-6: Valores de la función de puntuación de Rosetta obtenidos durante el *docking* molecular. * Rosetta Energy Units.

Los resultados derivados del *docking* molecular indican que se deberían priorizar los péptidos PLYQTWRFYF y KFMLRNERQF debido a que cuentan con los mejores valores en términos de

puentes de hidrógeno y “energía” del complejo péptido-HLA. En particular el péptido `PLYQTWRFYF` ya que éste cuenta con los mejores puntajes.

5. Conclusiones y perspectivas

5.1 Conclusiones

- En este trabajo se implementó una estrategia *in-silico* que permite identificar péptidos mutantes candidatos a vacuna personalizada en tumores cancerígenos a partir de la información genética de células tumorales y normales (exomas normal y tumoral y transcriptoma tumoral). Dicha estrategia se probó en el caso de una paciente con PEComa y que expresa el alelo HLA-A*24:02, dando como resultado la identificación de 12 péptidos candidatos a vacuna de los cuales 6 enmarcan una epítotope tanto para clase I como para clase II.
- Se implementó una estrategia *in-silico* para simular el *docking* molecular entre péptidos cortos (de 9 a 10 aminoácidos) y la proteína HLA-A*24:02, permitiendo identificar y evaluar características (puentes de hidrógeno y estimación de la energía del complejo) que pueden apoyar el proceso de priorización de péptidos mutantes cortos identificados como candidatos a vacuna personalizada en tumores cancerígenos.

5.2 Perspectivas

- El proceso de identificación de péptidos mutantes candidatos a vacuna personalizada en tumores cancerígenos está apoyado en un conjunto de herramientas *in-silico* que a la fecha cuentan con un considerable margen de error (tanto a nivel de falsos positivos como de falsos negativos) como se puede evidenciar en (3–6, 8, 12). Por tal motivo se hace necesario trabajar en la mejora de la precisión y exactitud de cada uno de los pasos involucrados en dicho proceso de identificación.
- El trabajo actual se concentró en implementar una estrategia *in-silico* de simulación del *docking* molecular para péptidos cortos, sería muy pertinente y valioso extender esta estrategia para péptidos largos (entre 15 y 21 aminoácidos) que son los que intervienen en la inmunogenicidad mediada por las moléculas HLA clase II y los linfocitos CD4+.
- Aunque el docking molecular aporta información relevante que puede ayudar en la toma de decisión a la hora de priorizar péptidos candidatos a vacuna personalizada para ser evaluados, por ejemplo, a nivel *in-vitro* o *in-vivo*. Los resultados que éste

arroja pueden ser aún mejor aprovechados si a partir de ellos se realizan análisis de dinámica molecular, ya que tal como se menciona en (14) la estabilidad del complejo péptido-HLA es una propiedad crítica en la inmunogenicidad de los péptidos. Aunque esta propiedad es considerada en el proceso de identificación implementado en el presente trabajo a través de la herramienta NetMHCStab (45), todavía tiene mucho margen de mejora por lo que la dinámica molecular podría realizar un aporte significativo en este aspecto.

- El presente trabajo se limitó a utilizar los datos genéticos (ADN y ARN) para la identificación de las mutaciones somáticas y cuantificar sus respectivos niveles de expresión. Sin embargo, es bien sabido que este tipo de datos son susceptibles de ser mejor explotados (por ejemplo a través de minería de datos o biología de sistemas) para obtener información e incluso conocimiento sobre el rol biológico que pueden desempeñar las proteínas que éstos codifican. Por lo cual se abre una ventana de oportunidad para integrar la información/conocimiento que se pueda extraer de estos datos genéticos a estrategias como la presentada en el actual trabajo.

6. Bibliografía

1. J. Hundal, B. M. Carreno, A. A. Petti, G. P. Linette, O. L. Griffith, E. R. Mardis, M. Griffith, pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. *Genome Med.* **8** (2016), doi:10.1186/s13073-016-0264-5.
2. A. Rubinsteyn, J. Kodysh, I. Hodes, S. Mondet, B. A. Aksoy, J. P. Finnigan, N. Bhardwaj, J. Hammerbacher, Computational pipeline for the PGV-001 neoantigen vaccine trial. *Front. Immunol.* **8**, 1807 (2018).
3. M. Yadav, S. Jhunjunwala, Q. T. Phung, P. Lupardus, J. Tanguay, S. Bumbaca, C. Franci, T. K. Cheung, J. Fritsche, T. Weinschenk, Z. Modrusan, I. Mellman, J. R. Lill, L. Delamarre, Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature.* **515**, 572–576 (2014).
4. M. M. Gubin, X. Zhang, H. Schuster, E. Caron, J. P. Ward, T. Noguchi, Y. Ivanova, J. Hundal, C. D. Arthur, W.-J. Krebber, G. E. Mulder, M. Toebes, M. D. Vesely, S. S. K. Lam, A. J. Korman, J. P. Allison, G. J. Freeman, A. H. Sharpe, E. L. Pearce, T. N. Schumacher, R. Aebbersold, H.-G. Rammensee, C. J. M. Melief, E. R. Mardis, W. E. Gillanders, M. N. Artyomov, R. D. Schreiber, Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature.* **515**, 577–581 (2014).
5. C. Linnemann, M. M. van Buuren, L. Bies, E. M. E. Verdegaal, R. Schotte, J. J. A. Calis, S. Behjati, A. Velds, H. Hilkmann, D. el Atmioui, M. Visser, M. R. Stratton, J. B. A. G. Haanen, H. Spits, S. H. van der Burg, T. N. M. Schumacher, High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4⁺ T cells in human melanoma. *Nat. Med.* **21**, 81–85 (2015).
6. B. M. Carreno, V. Magrini, M. Becker-Hapak, S. Kaabinejadian, J. Hundal, A. A. Petti, A. Ly, W.-R. Lie, W. H. Hildebrand, E. R. Mardis, G. P. Linette, A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science.* **348**, 803–808 (2015).
7. N. A. Rizvi, M. D. Hellmann, A. Snyder, P. Kvistborg, V. Makarov, J. J. Havel, W. Lee, J. Yuan, P. Wong, T. S. Ho, M. L. Miller, N. Rekhtman, A. L. Moreira, F. Ibrahim, C. Bruggeman, B. Gasmi, R. Zappasodi, Y. Maeda, C. Sander, E. B. Garon, T. Merghoub, J. D. Wolchok, T. N. Schumacher, T. A. Chan, Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* **348**, 124–128 (2015).
8. X. Zhang, S. Kim, J. Hundal, J. M. Herndon, S. Li, A. A. Petti, S. D. Soysal, L. Li, M. D. McLellan, J. Hoog, T. Primeau, N. Myers, T. L. Vickery, M. Sturmoski, I. S. Hagemann, C. A. Miller, M. J. Ellis, E. R. Mardis, T. Hansen, T. P. Fleming, S. P. Goedegebuure, W. E. Gillanders, Breast Cancer Neoantigens Can Induce CD8⁺ T-Cell Responses and Antitumor Immunity. *Cancer Immunol. Res.* **5**, 516–523 (2017).
9. U. Sahin, Ö. Türeci, Personalized vaccines for cancer immunotherapy. *Science.* **359**, 1355–1360 (2018).
10. V. Jurtz, S. Paul, M. Andreatta, P. Marcatili, B. Peters, M. Nielsen, NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J. Immunol.* **199**, 3360–3368 (2017).
11. T. J. O'Donnell, A. Rubinsteyn, M. Bonsack, A. B. Riemer, U. Laserson, J. Hammerbacher,

- MHCflurry: Open-Source Class I MHC Binding Affinity Prediction. *Cell Syst.* **7**, 129-132.e4 (2018).
12. W. Zhao, X. Sher, Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes. *PLoS Comput. Biol.* **14**, e1006457 (2018).
 13. M. Harndahl, M. Rasmussen, G. Roder, I. Dalgaard Pedersen, M. Sørensen, M. Nielsen, S. Buus, Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. *Eur. J. Immunol.* **42**, 1405–1416 (2012).
 14. The problem with neoantigen prediction. *Nat. Biotechnol.* **35**, 97 (2017).
 15. Global Cancer Observatory, (available at <http://gco.iarc.fr/>).
 16. World Health Organization, *World Health Statistics 2018: monitoring health for the SDGs : sustainable development goals.* (2018).
 17. Cancer today - Global Cancer Observatory, (available at <http://gco.iarc.fr/today/home>).
 18. M. H. Manjili, Revisiting cancer immunoediting by understanding cancer immune complexity. *J. Pathol.* **224**, 5–9 (2011).
 19. S. H. Hassanpour, M. Dehghani, Review of cancer from perspective of molecular. *J. Cancer Res. Pract.* **4**, 127–129 (2017).
 20. D. Mittal, M. M. Gubin, R. D. Schreiber, M. J. Smyth, New insights into cancer immunoediting and its three component phases — elimination, equilibrium and escape. *Curr. Opin. Immunol.* **27**, 16–25 (2014).
 21. R. D. Wood, M. Mitchell, J. Sgouros, T. Lindahl, Human DNA Repair Genes. *Science.* **291**, 1284–1289 (2001).
 22. G. Martignoni, M. Pea, D. Reghellin, G. Zamboni, F. Bonetti, PEComas: the past, the present and the future. *Virchows Arch.* **452**, 119–132 (2008).
 23. Apitz K, Die Geschwülste und Gewebsmissbildungen der Nierenrinde. *Virchows Arch.* **311**, 306–327 (1944).
 24. M. Pea, F. Bonetti, G. Zamboni, G. Martignoni, L. Fiore-Donati, C. Doglioni, Clear cell tumor and angiomyolipoma. *Am. J. Surg. Pathol.* **15**, 199–200 (1991).
 25. F. Bonetti, M. Pea, G. Martignoni, G. Zamboni, PEC and sugar. *Am. J. Surg. Pathol.* **16**, 307 (1992).
 26. J. S. Bleeker, J. F. Quevedo, A. L. Folpe, “Malignant” perivascular epithelioid cell neoplasm: risk stratification and treatment strategies. *Sarcoma.* **2012** (2012), doi:<https://doi.org/10.1155/2012/541626>.
 27. A. Zimmermann, in *Tumors and Tumor-Like Lesions of the Hepatobiliary Tract: General and Surgical Pathology* (Springer, 2016; https://link.springer.com/referenceworkentry/10.1007%2F978-3-319-26587-2_73-1).
 28. F. Bonetti, M. Pen, G. Martignoni, G. Zamboni, E. Manirin, R. Colombari, G. M. Mariuzzi, The perivascular epithelioid cell and related lesions. *Adv. Anat. Pathol.* **4**, 343–358 (1997).
 29. A. K. Abbas, A. H. H. Lichtman, S. Pillai, *Inmunología celular y molecular* (Elsevier Health Sciences Spain, 2015).
 30. Parra Lopez, Carlos, Curso: Fundamentos de Inmunología: Conceptos de Procesamiento y Presentación del Antígeno. *Univ. Nac. Colomb.* (2013).
 31. Delgado Murcia G., Curso “La nueva Inmunología Molecular, Generalidades del Procesamiento y Presentación del Antígeno.” *Fund. Inst. Inmunol. Colomb.* (2002).
 32. S. Tenzer, B. Peters, S. Bulik, O. Schoor, C. Lemmel, M. M. Schatz, P.-M. Kloetzel, H.-G. Rammensee, H. Schild, H.-G. Holzhütter, Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell. Mol. Life Sci. CMLS.* **62**, 1025–1037 (2005).
 33. S. El-Metwally, T. Hamza, M. Zakaria, M. Helmy, Next-generation sequence assembly: four stages of data processing and computational challenges. *PLoS Comput. Biol.* **9**, e1003345 (2013).
 34. E. C. Hayden, Is the \$1,000 genome for real? *Nat. News* (2014).

35. D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, Accurate whole human genome sequencing using reversible terminator chemistry. *nature*. **456**, 53 (2008).
36. S. W. Kowalczyk, D. B. Wells, A. Aksimentiev, C. Dekker, Slowing down DNA translocation through a nanopore in lithium chloride. *Nano Lett.* **12**, 1038–1044 (2012).
37. H. Esfandyarpour, Genapsys 100X Solution: Label-free Fully-integrated “Personal Genomixer.” *J. Biomol. Tech. JBT*. **23**, S9 (2012).
38. K. Paszkiewicz, D. J. Studholme, De novo assembly of short sequence reads. *Brief. Bioinform.* **11**, 457–472 (2010).
39. G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, *Curr. Protoc. Bioinforma.*, in press.
40. A. N. Houghton, J. A. Guevara-Patiño, Immune recognition of self in immunity against cancer. *J. Clin. Invest.* **114**, 468–471 (2004).
41. H. Zhang, O. Lund, M. Nielsen, The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. *Bioinformatics*. **25**, 1293–1299 (2009).
42. M. Andreatta, M. Nielsen, Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*. **32**, 511–517 (2015).
43. E. Karosiene, C. Lundegaard, O. Lund, M. Nielsen, NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics*. **64**, 177–186 (2012).
44. M. Nielsen, C. Lundegaard, O. Lund, C. Keşmir, The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*. **57**, 33–41 (2005).
45. K. W. Jørgensen, M. Rasmussen, S. Buus, M. Nielsen, Net MHC stab–predicting stability of peptide–MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology*. **141**, 18–26 (2014).
46. W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The Ensembl Variant Effect Predictor. *Genome Biol.* **17** (2016), doi:10.1186/s13059-016-0974-4.
47. A. Szolek, B. Schubert, C. Mohr, M. Sturm, M. Feldhahn, O. Kohlbacher, OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics*. **30**, 3310–3316 (2014).
48. S. Boegel, M. Löwer, M. Schäfer, T. Bukur, J. de Graaf, V. Boisguérin, Ö. Türeci, M. Diken, J. C. Castle, U. Sahin, HLA typing from RNA-Seq sequence reads. *Genome Med.* **4**, 102 (2012).
49. K. K. Jensen, M. Andreatta, P. Marcatili, S. Buus, J. A. Greenbaum, Z. Yan, A. Sette, B. Peters, M. Nielsen, Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*. **154**, 394–406 (2018).
50. Y. Kim, J. Sidney, C. Pinilla, A. Sette, B. Peters, Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. *BMC Bioinformatics*. **10**, 394 (2009).
51. R. Bhattacharya, C. Tokheim, A. Sivakumar, V. B. Guthrie, V. Anagnostou, V. E. Velculescu, R. Karchin, Prediction of peptide binding to MHC Class I proteins in the age of deep learning. *bioRxiv*, 154757 (2017).
52. X. M. Shao, R. Bhattacharya, J. Huang, I. A. Sivakumar, C. Tokheim, L. Zheng, D. Hirsch, B. Kaminow, A. Omdahl, M. Bonsack, High-throughput prediction of MHC Class I and Class II neoantigens with MHCnuggets. *bioRxiv*, 752469 (2019).
53. M. Nielsen, C. Lundegaard, O. Lund, Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics*. **8**, 238

- (2007).
54. M. Fermeglia, S. Pricl, G. Longo, Molecular modeling and process simulation: Real possibilities and challenges. *Chem. Biochem. Eng. Q.* **17**, 19–30 (2003).
 55. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
 56. Y. Zhang, I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* **9**, 40 (2008).
 57. W. J. Hehre, *A guide to molecular mechanics and quantum chemical calculations* (Wavefunction Irvine, CA, 2003), vol. 2.
 58. Schrodinger, The PyMOL Molecular Graphics System, Version 1.8 (2015).
 59. U. Chimera, a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J Comput Chem.* **25**, 1605–12 (2004).
 60. X.-Y. Meng, H.-X. Zhang, M. Mezei, M. Cui, Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug Des.* **7**, 146–157 (2011).
 61. O. Schueler-Furman, N. London, Schueler-Furman, *Modeling peptide-protein interactions* (Springer, 2017).
 62. B. Raveh, N. London, L. Zimmerman, O. Schueler-Furman, Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. *PloS One.* **6**, e18934 (2011).
 63. B. Raveh, N. London, O. Schueler-Furman, Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins Struct. Funct. Bioinforma.* **78**, 2029–2040 (2010).
 64. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, in *Methods in enzymology* (Elsevier, 2011), vol. 487, pp. 545–574.
 65. R. Das, D. Baker, Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **77**,. **363** (2008).
 66. K. W. Kaufmann, G. H. Lemmon, S. L. DeLuca, J. H. Sheehan, J. Meiler, Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry.* **49**, 2987–2998 (2010).
 67. F. Poy, M. B. Yaffe, J. Sayos, K. Saxena, M. Morra, J. Sumegi, L. C. Cantley, C. Terhorst, M. J. Eck, Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. *Mol. Cell.* **4**, 555–561 (1999).
 68. Y. Li, K. Suino, J. Daugherty, H. E. Xu, Structural and biochemical mechanisms for the specificity of hormone binding and coactivator assembly by mineralocorticoid receptor. *Mol. Cell.* **19**, 367–380 (2005).
 69. S. Andrews, *FastQC: a quality control tool for high throughput sequence data* (Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, 2010).
 70. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools. *Bioinformatics.* **25**, 2078–2079 (2009).
 71. T. Picard, *Broad Institute, GitHub repository* (2018).
 72. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, The variant call format and VCFtools. *Bioinformatics.* **27**, 2156–2158 (2011).
 73. O. Tange, Gnu parallel-the command-line power tool. *USENIX Mag.* **36**, 42–47 (2011).
 74. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* **25**, 1754–1760 (2009).
 75. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: accurate

alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

76. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
77. C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).

Anexo 1: Resultados del control de calidad realizado con FastQC

Exoma normal:

Estadísticas generales:

Lecturas <i>Forward</i>		Lecturas <i>Reverse</i>	
Magnitud	Valor	Magnitud	Valor
Lecturas totales	70'139.549	Lecturas totales	70'139.549
Lecturas de baja calidad	0	Lecturas de baja calidad	0
Longitud lectura	150	Longitud lectura	150
% GC	50	% GC	50

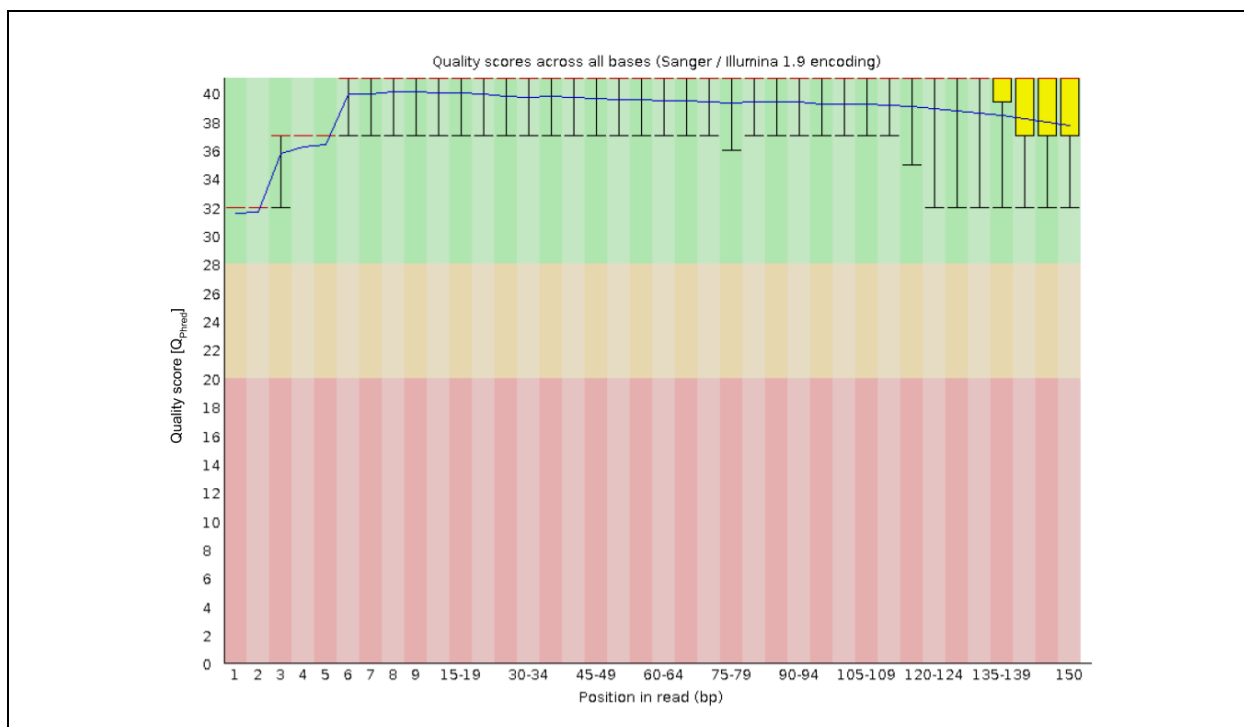


Figura S1-1: Puntajes promedio de calidad [Q_{Phred}] por posición en la lectura para el exoma normal (lecturas *Forward*).

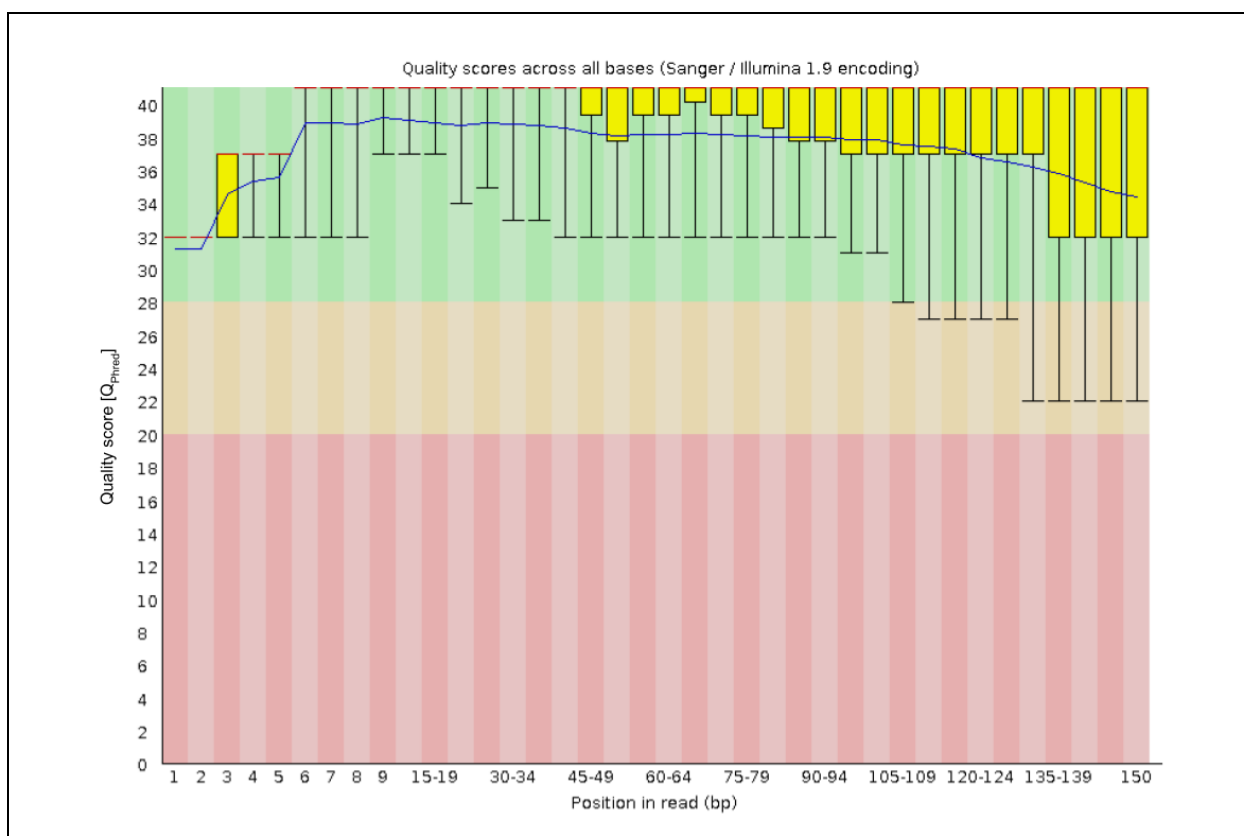


Figura S1-2: Puntajes promedio de calidad [Q_{phred}] por posición en la lectura para el exoma normal (lecturas *Reverse*).

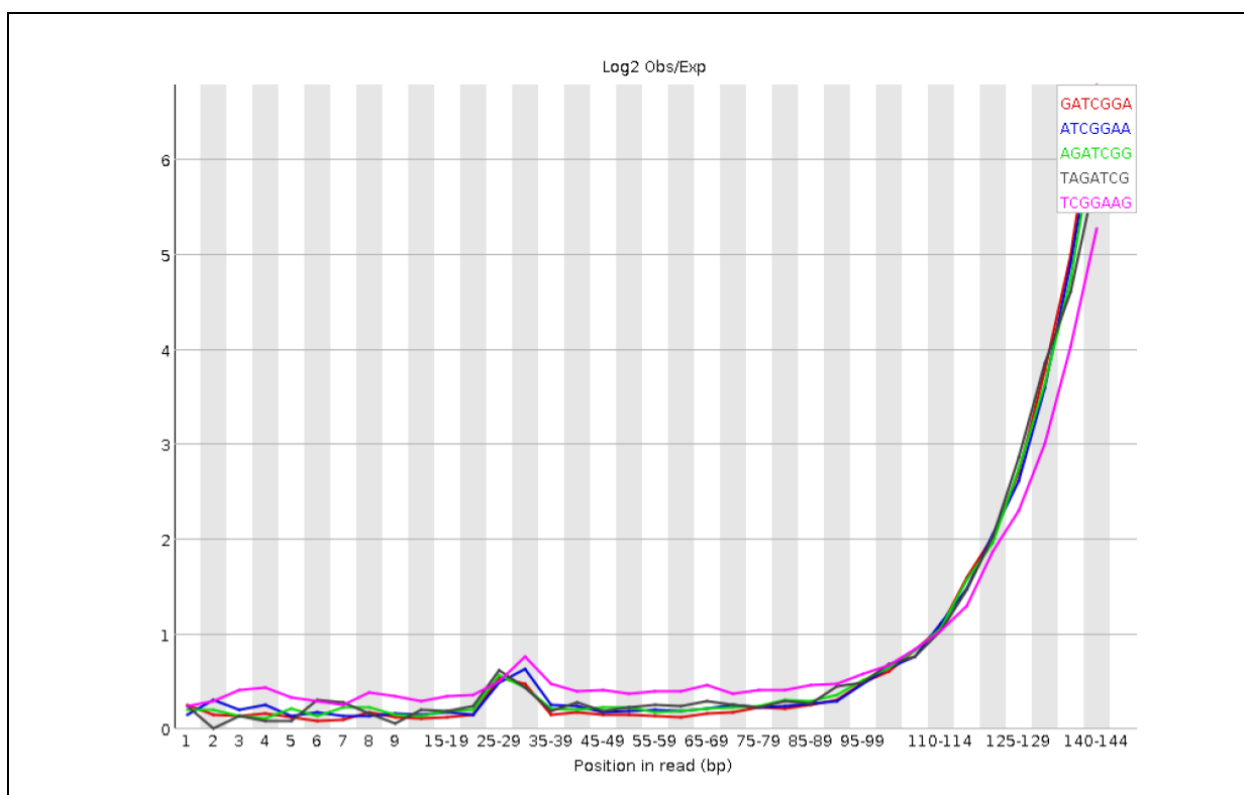


Figura S1-3: Distribución de *Kmers* por posición en la lectura para el exoma normal (lecturas *Forward*).

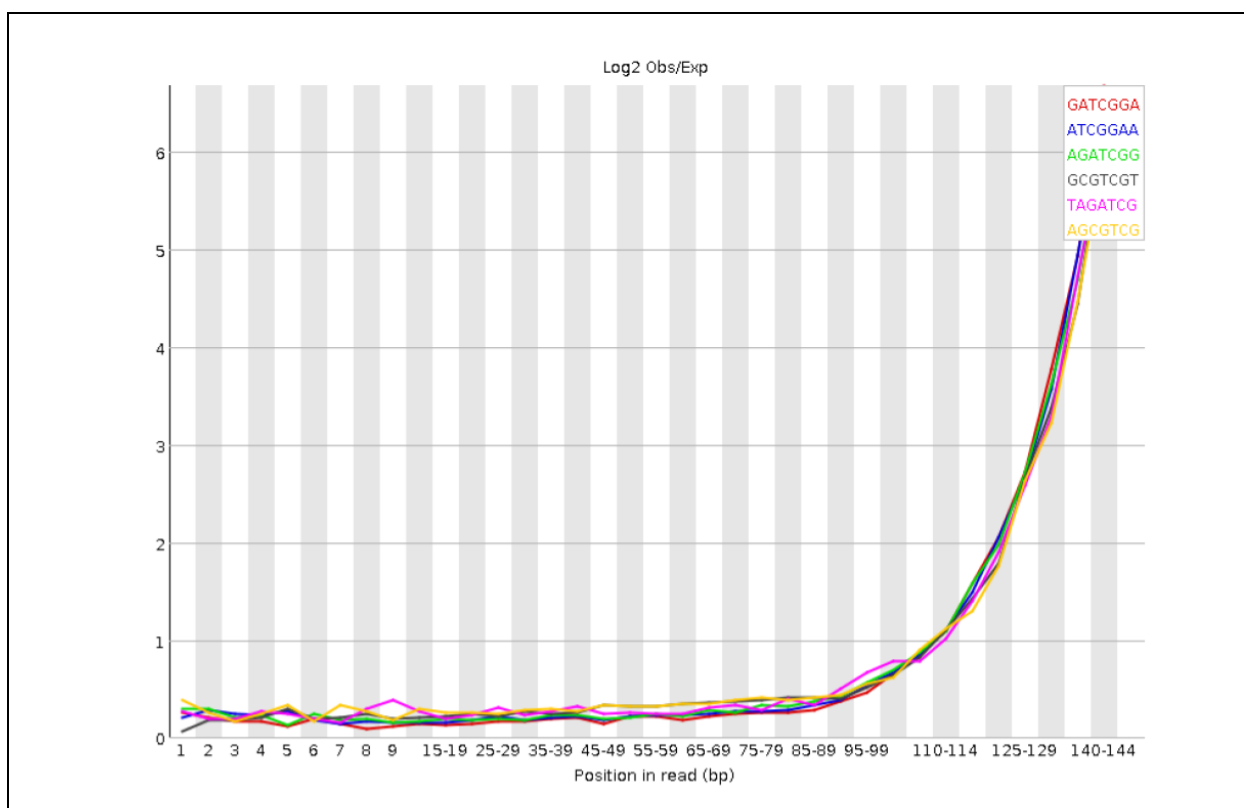


Figura S1-4: Distribución de *Kmers* por posición en la lectura para el exoma normal (lecturas *Reverse*).

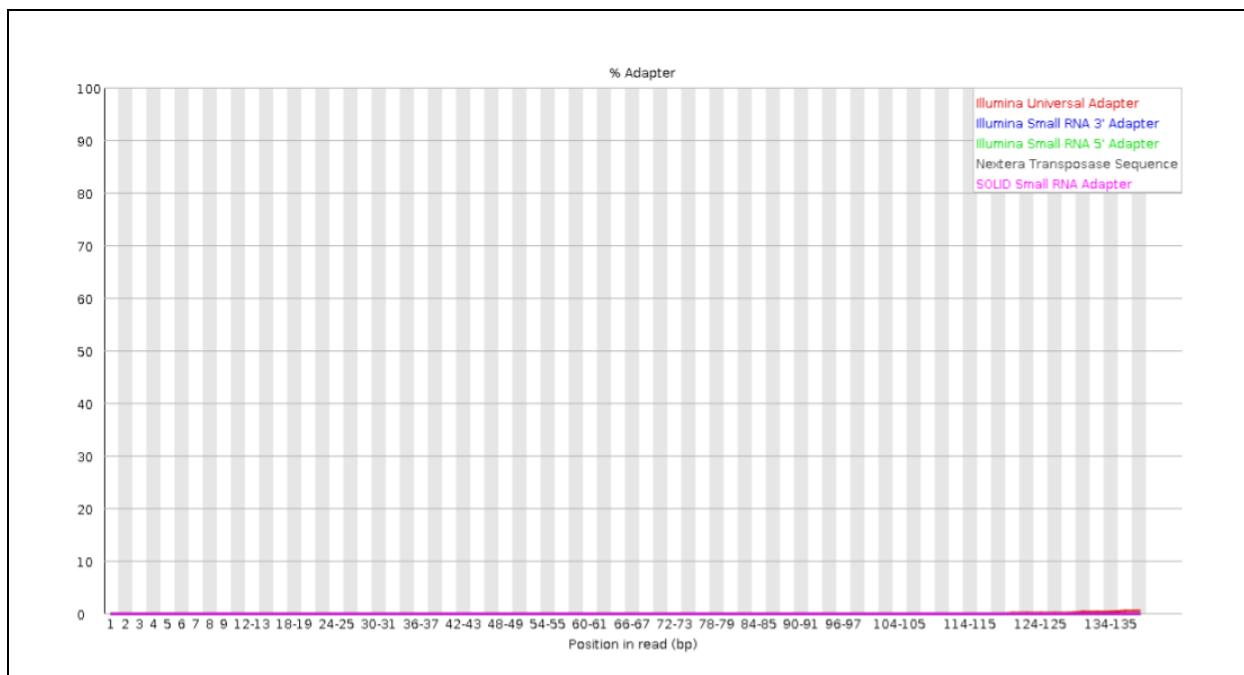


Figura S1-5: Distribución del contenido de adaptadores por posición en la lectura para el exoma normal (lecturas *Forward*).

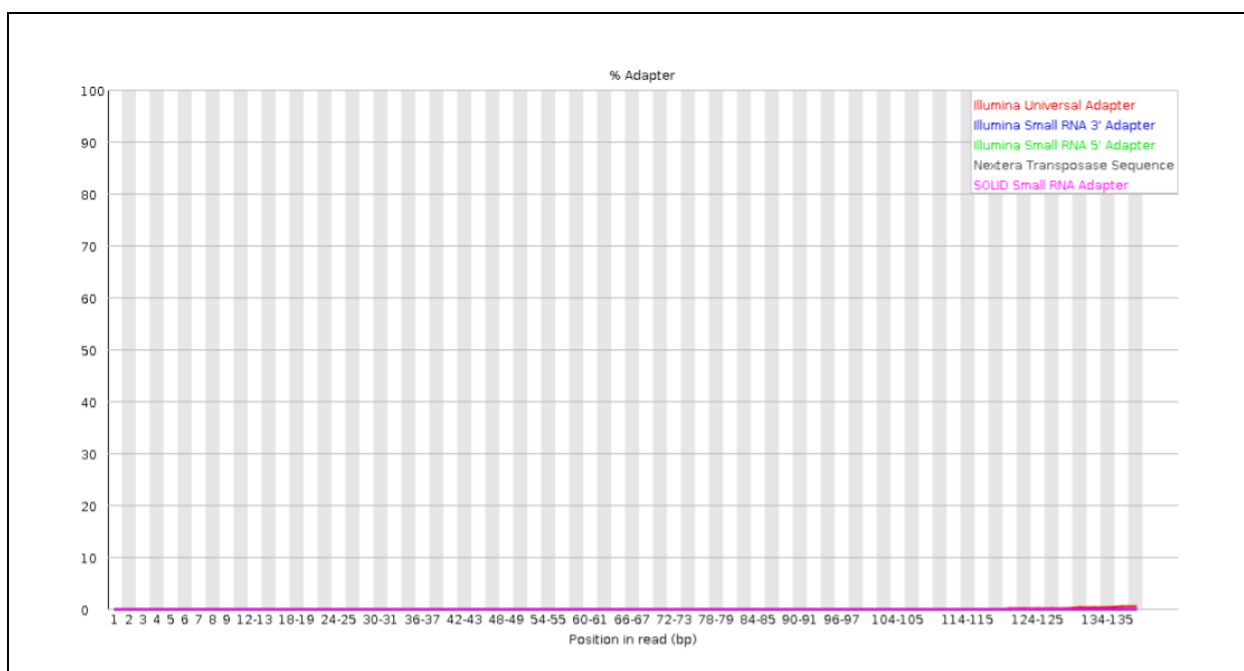


Figura S1-6: Distribución del contenido de adaptadores por posición en la lectura para el exoma normal (lecturas *Reverse*).

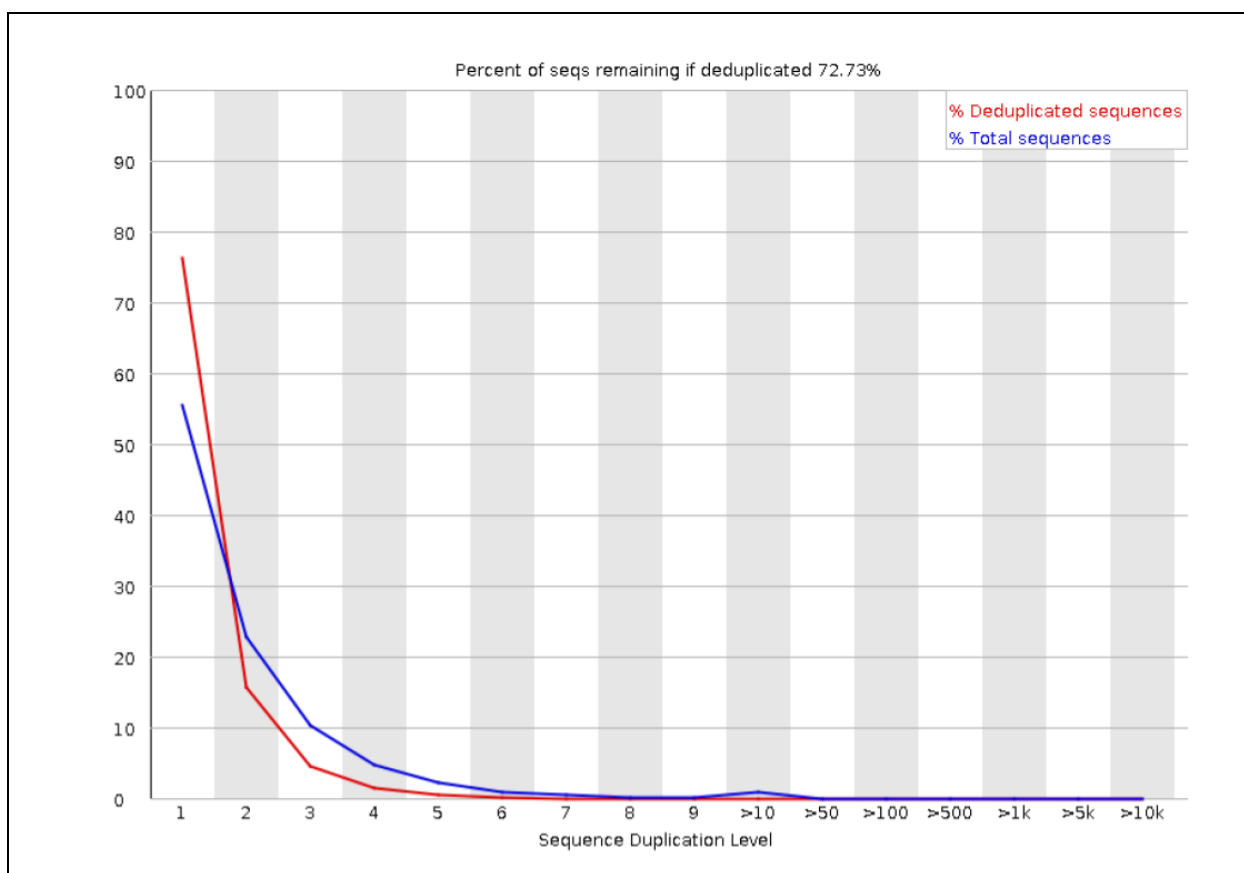


Figura S1-7: Distribución de los porcentajes de lecturas duplicadas de acuerdo al nivel de duplicación para el exoma normal (lecturas *Forward*).

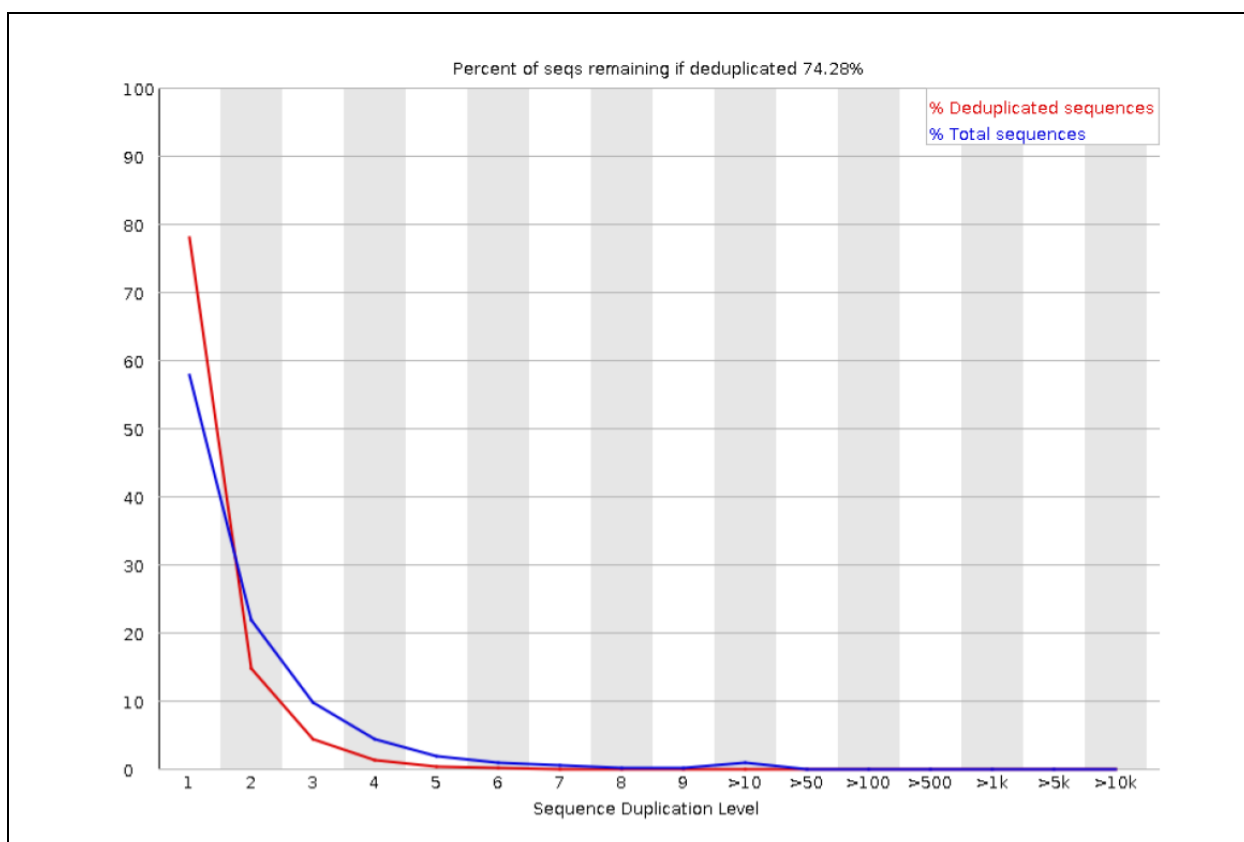


Figura S1-8: Distribución de los porcentajes de lecturas duplicadas de acuerdo al nivel de duplicación para el exoma normal (lecturas *Reverse*).

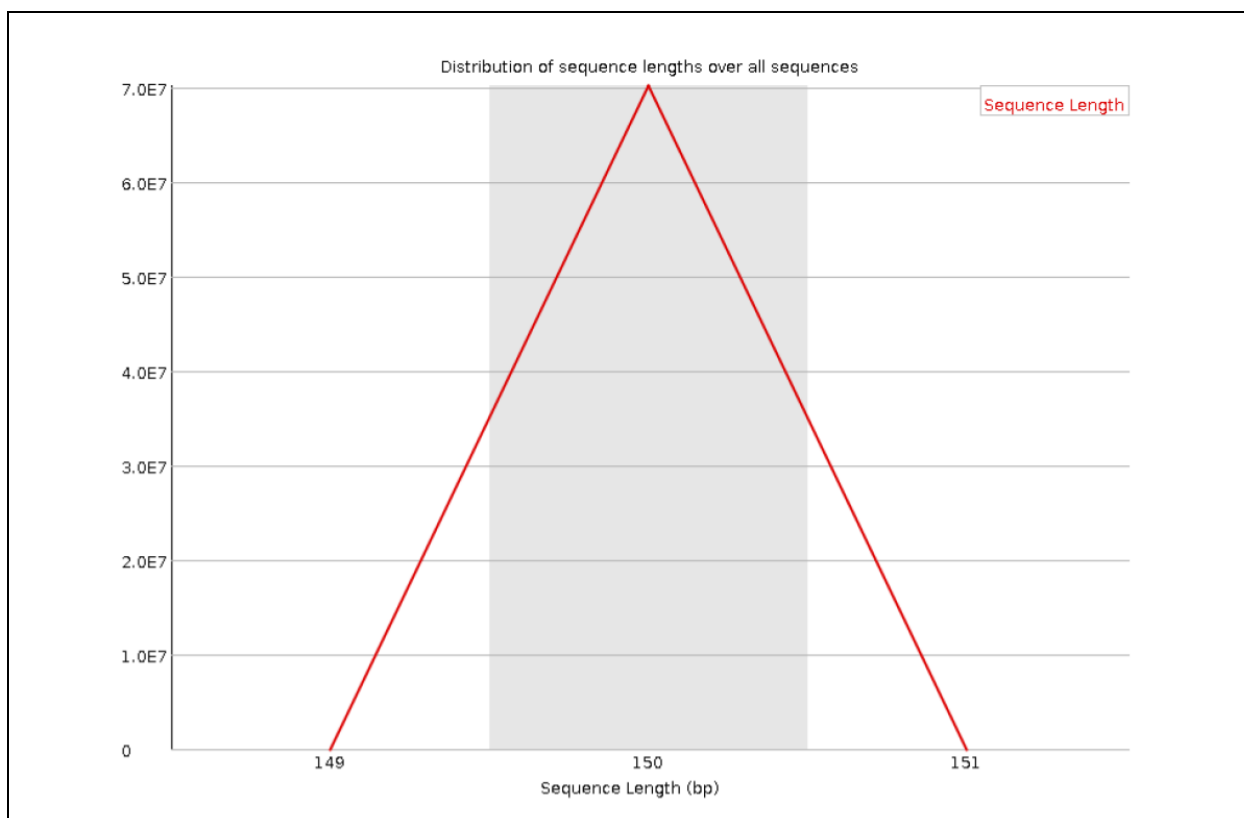


Figura S1-9: Distribución de las longitudes de lecturas para el exoma normal (lecturas *Forward*).

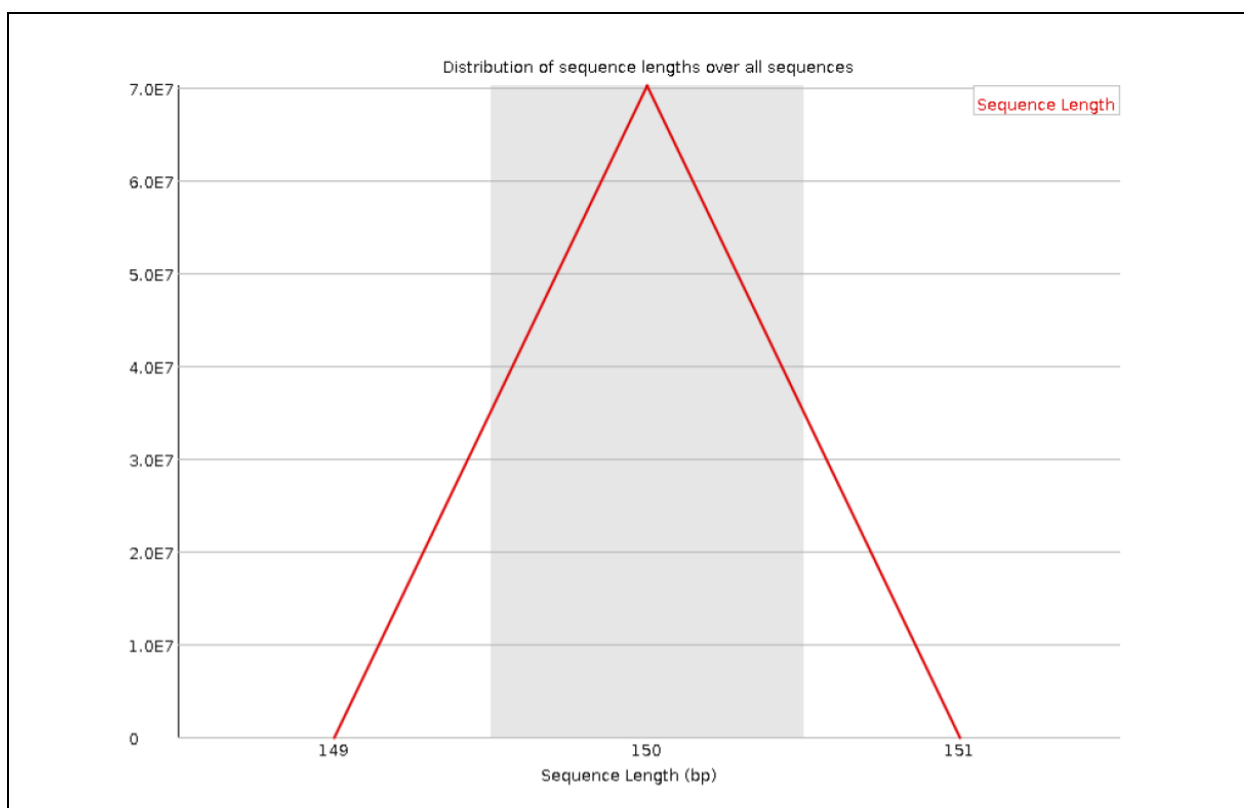


Figura S1-10: Distribución de las longitudes de lecturas para el exoma normal (lecturas *Reverse*).

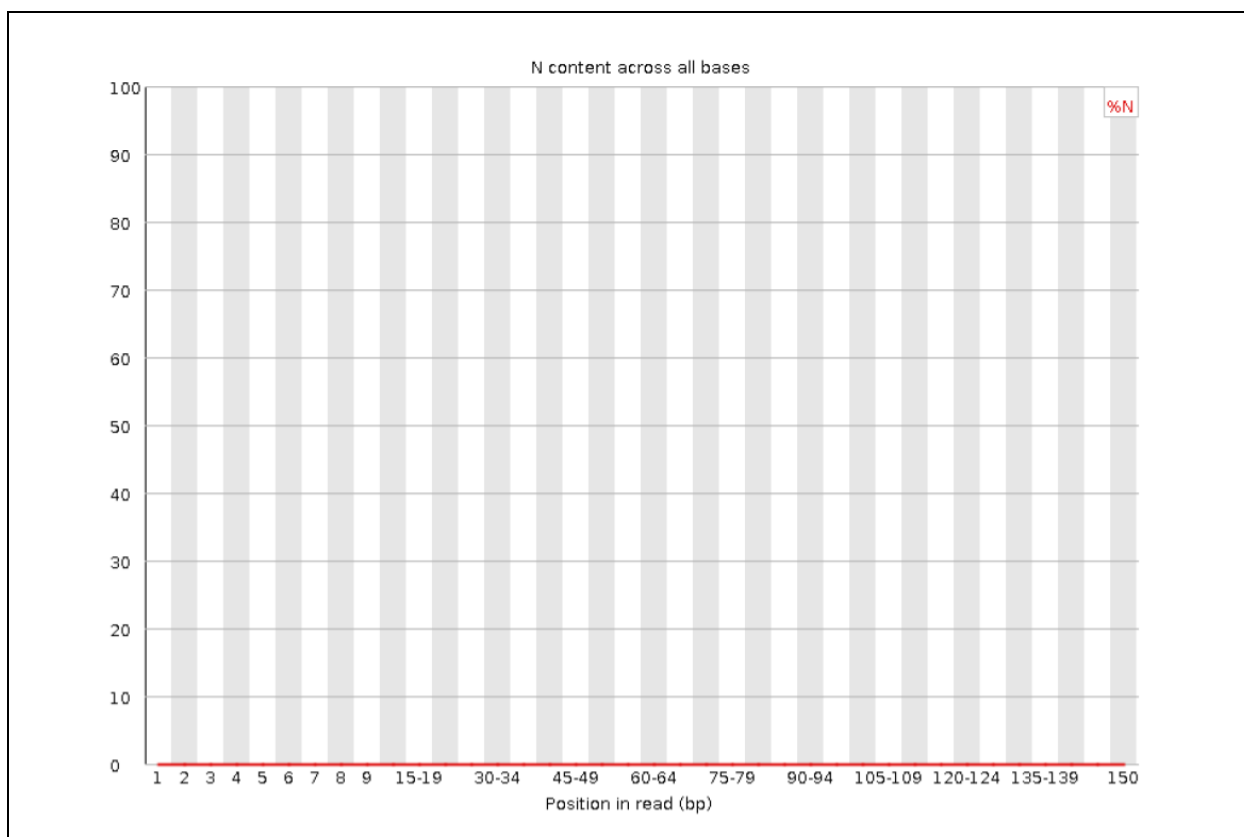


Figura S1-11: Distribución de los porcentajes de bases no identificadas por posición en la lectura para el exoma normal (lecturas *Forward*).

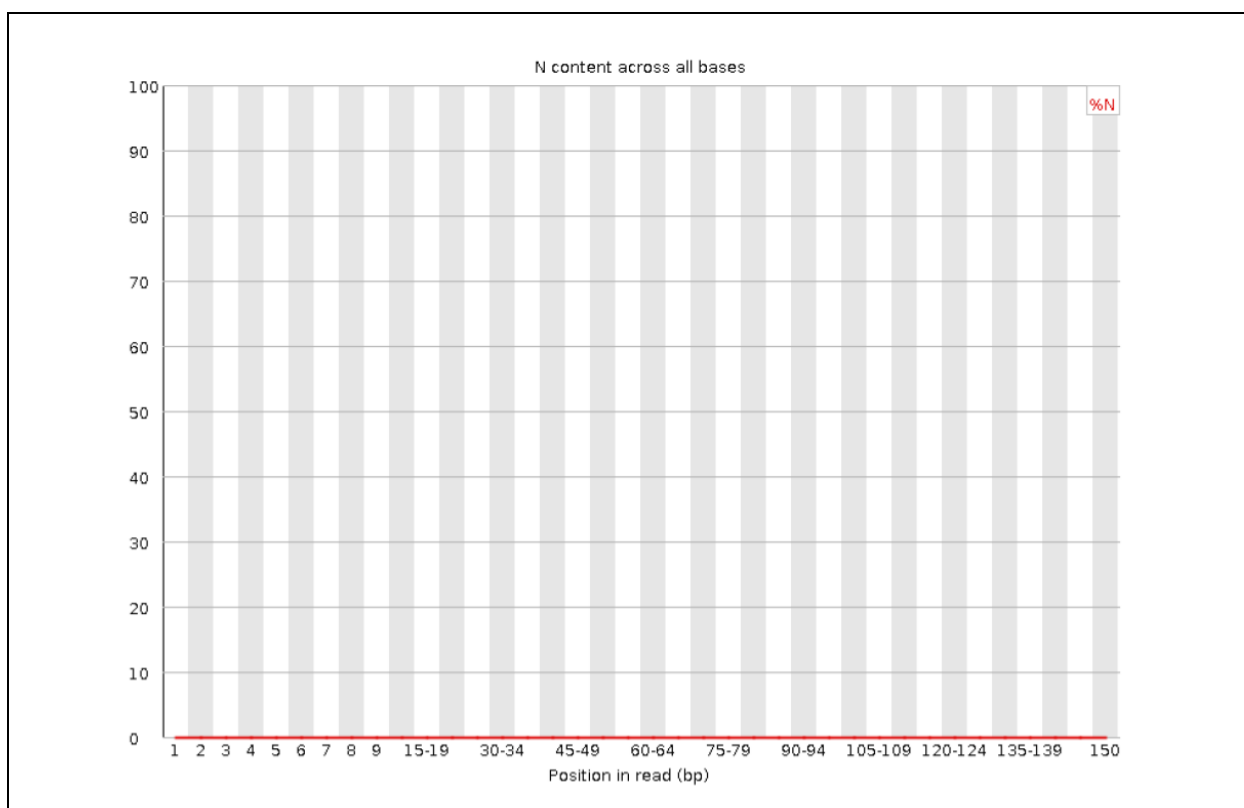


Figura S1-12: Distribución de los porcentajes de bases no identificadas por posición en la lectura para el exoma normal (lecturas *Reverse*).

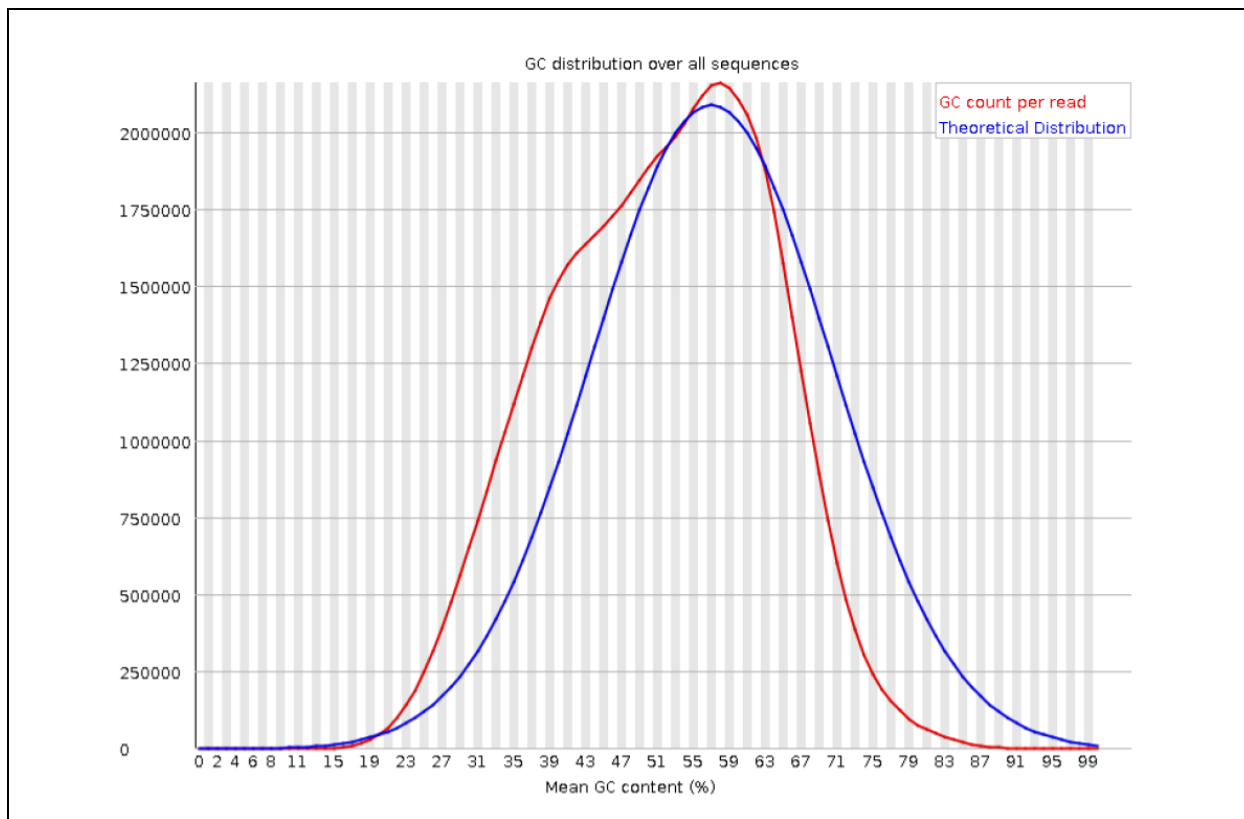


Figura S1-13: Distribución del contenido medio de GC en las lecturas para el exoma normal (lecturas *Forward*).

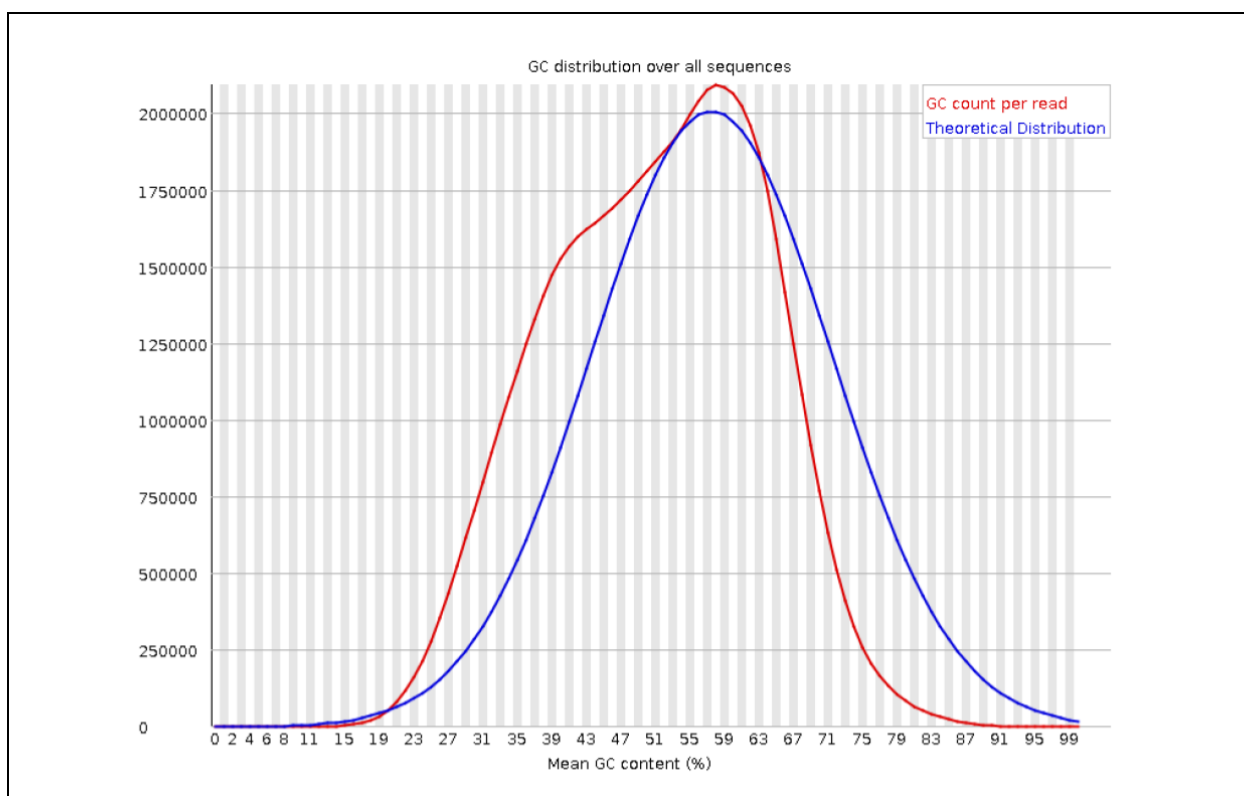


Figura S1-14: Distribución del contenido medio de GC en las lecturas para el exoma normal (lecturas *Reverse*).

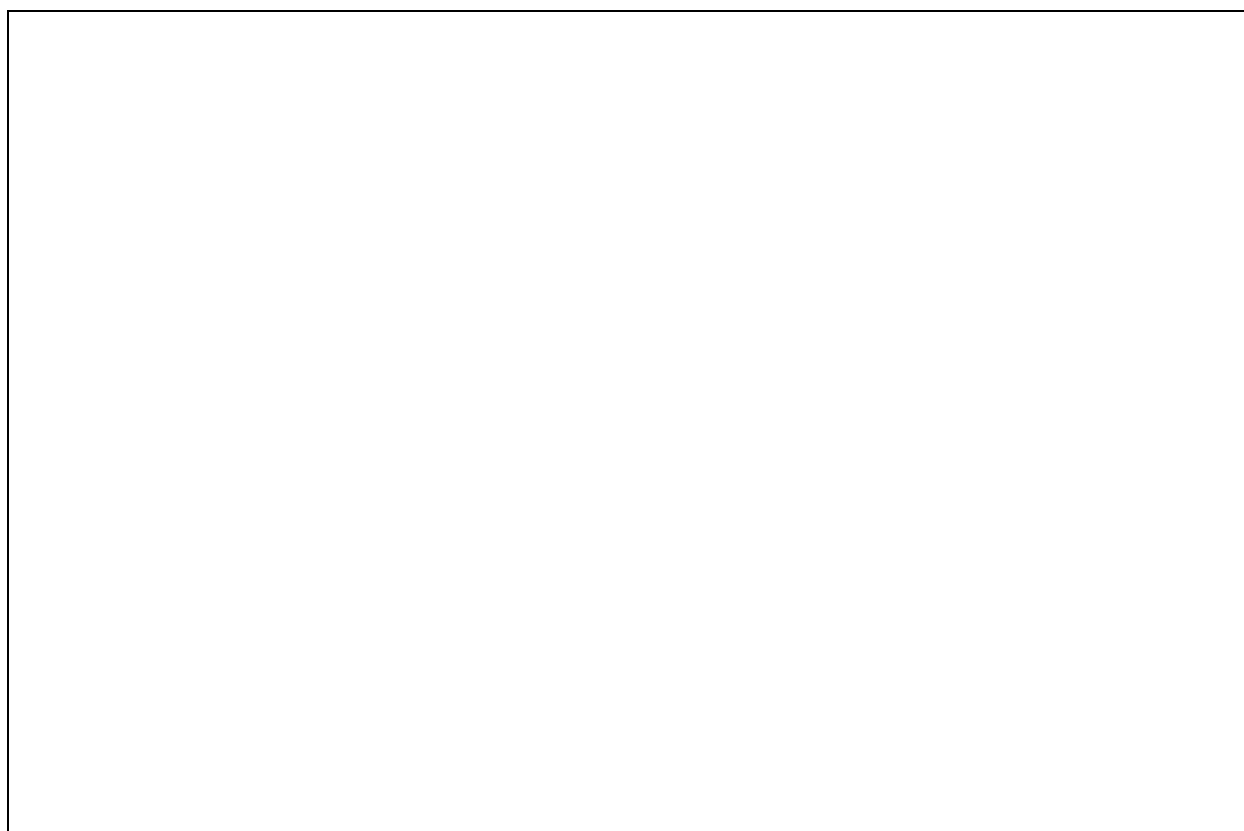


Figura S1-15: Distribución del contenido medio de GC por posición en la lectura para el exoma normal (lecturas *Forward*).

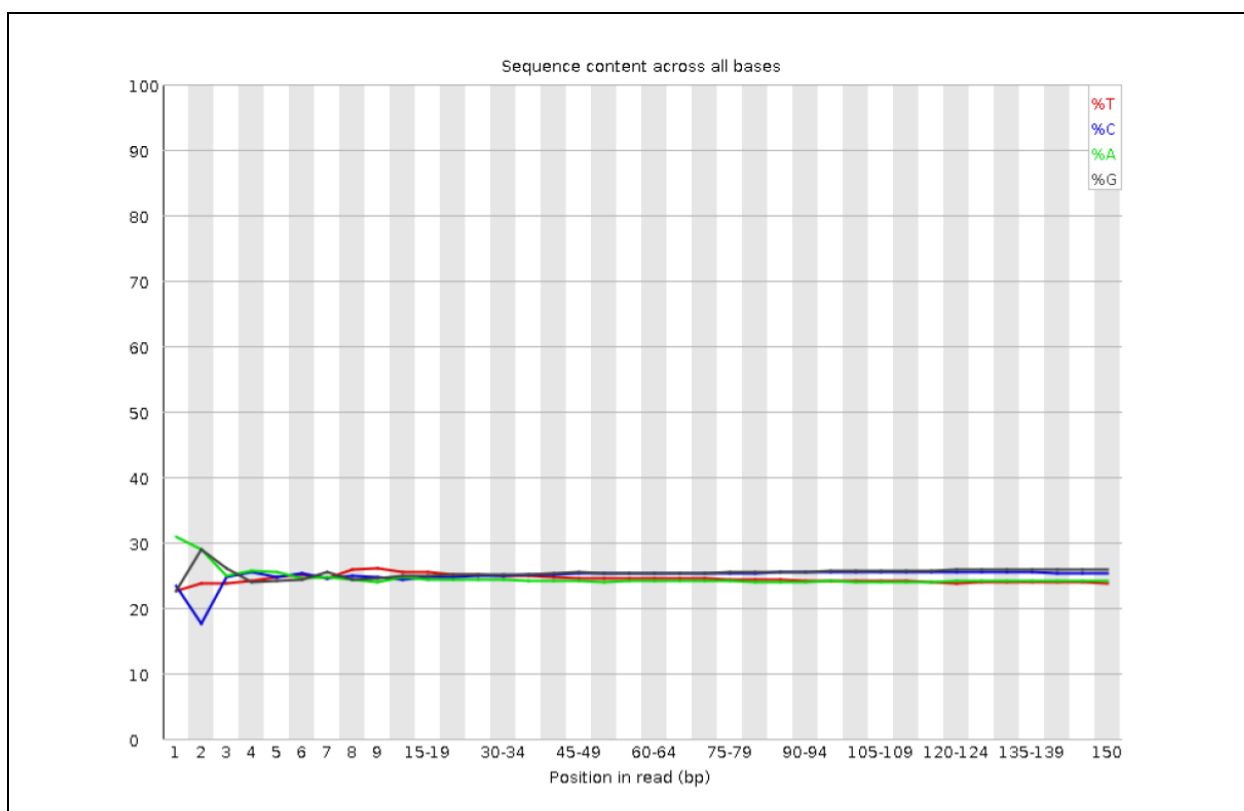


Figura S1-16: Distribución del contenido medio de GC por posición en la lectura para el exoma normal (lecturas *Reverse*).

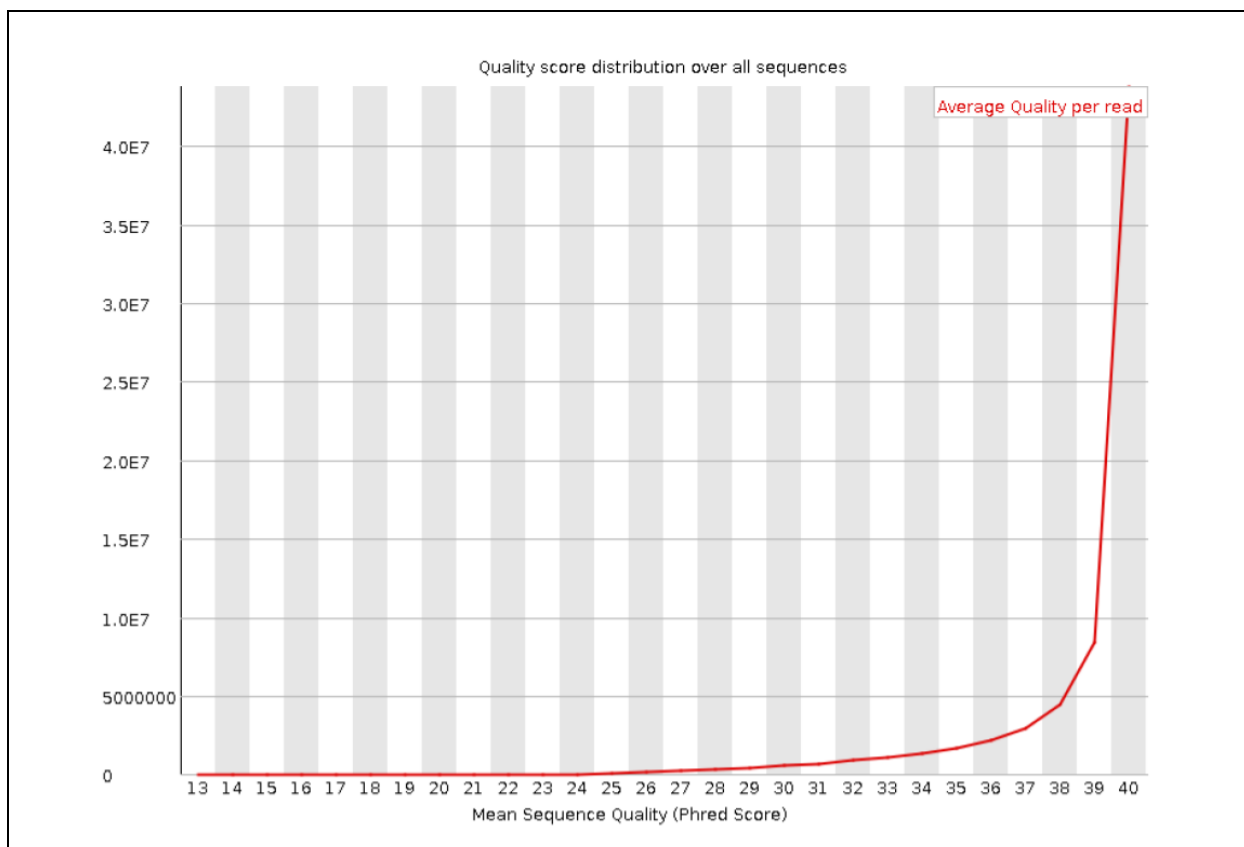


Figura S1-17: Distribución de los puntajes promedio de calidad por lectura para el exoma normal (lecturas *Forward*).

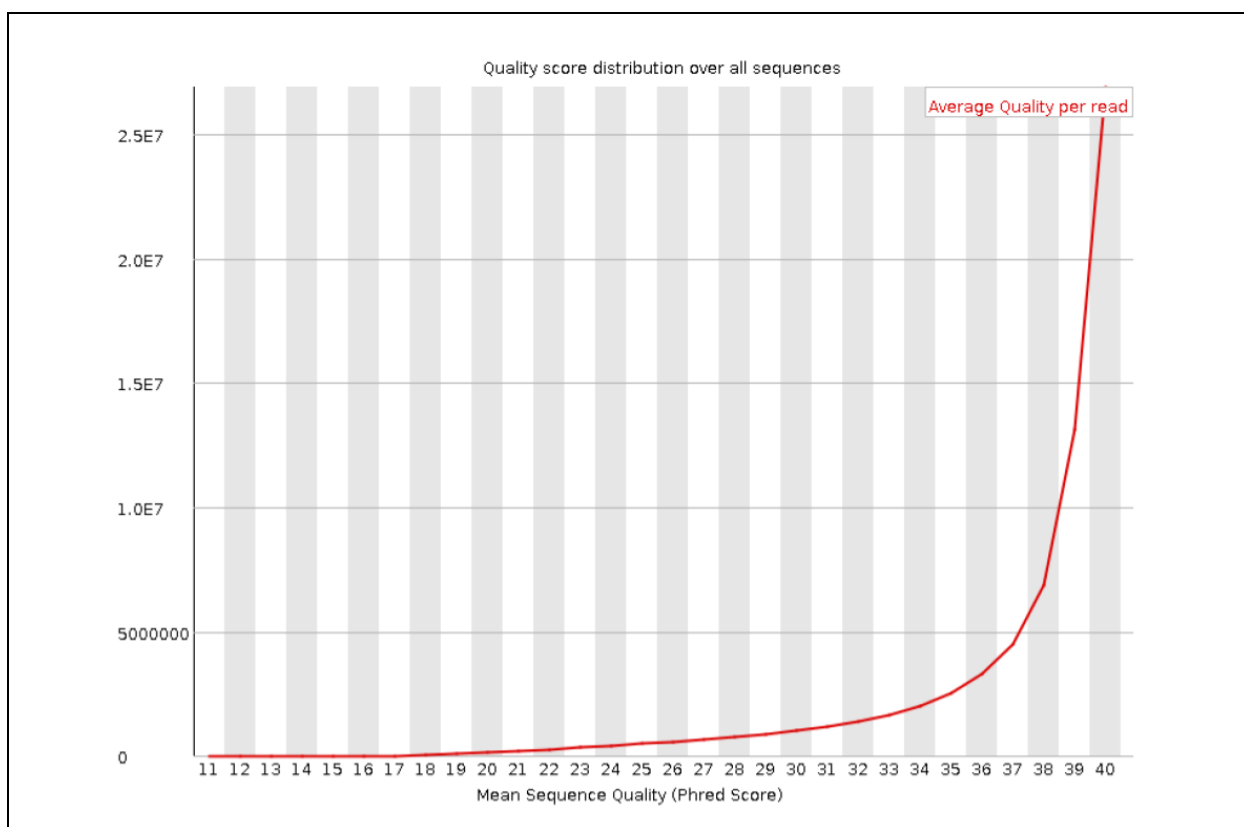


Figura S1-18: Distribución de los puntajes promedio de calidad por lectura para el exoma normal (lecturas *Reverse*).

Exoma tumoral:

Estadísticas generales:

Lecturas <i>Forward</i>		Lecturas <i>Reverse</i>	
Magnitud	Valor	Magnitud	Valor
Lecturas totales	120'041.101	Lecturas totales	120'041.101
Lecturas de baja calidad	0	Lecturas de baja calidad	0
Longitud lectura	150	Longitud lectura	150
% GC	50	% GC	50

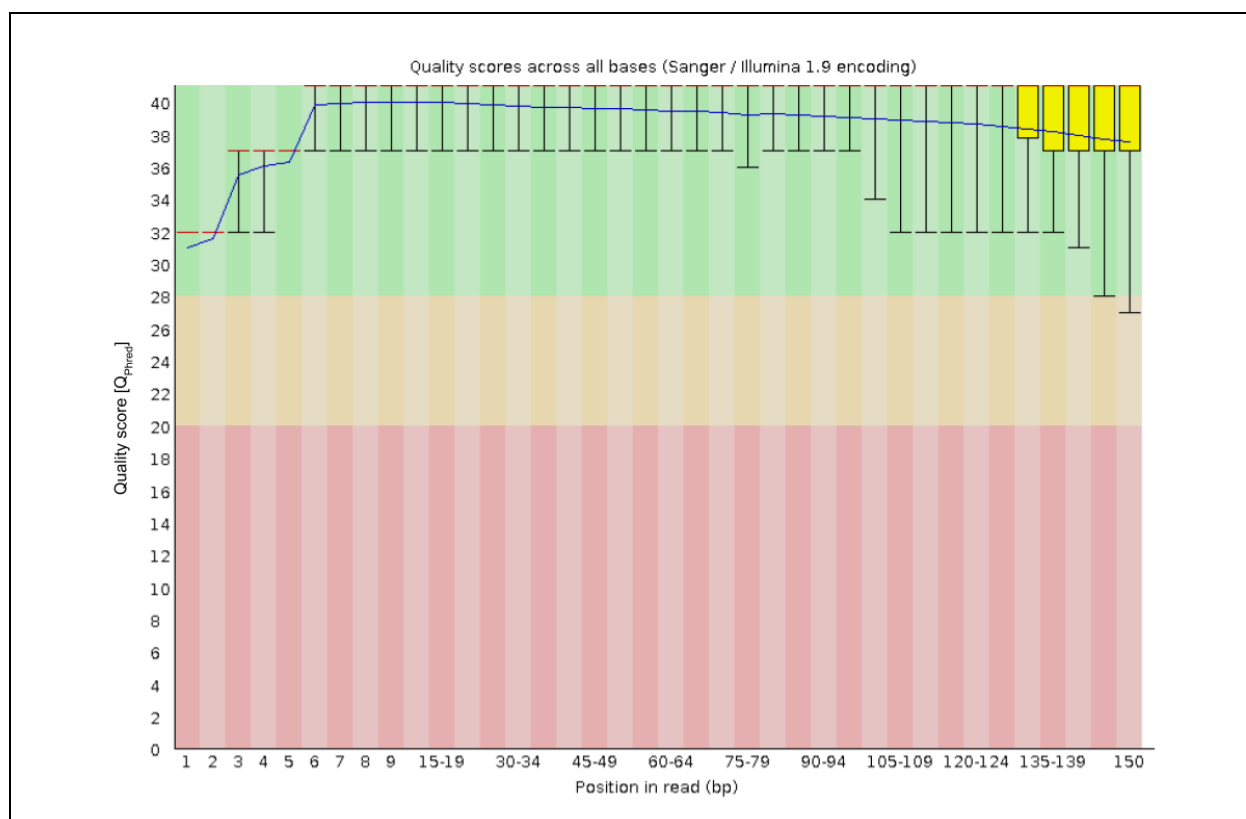


Figura S1-19: Puntajes promedio de calidad [Q_{phred}] por posición en la lectura para el exoma tumoral (lecturas *Forward*).

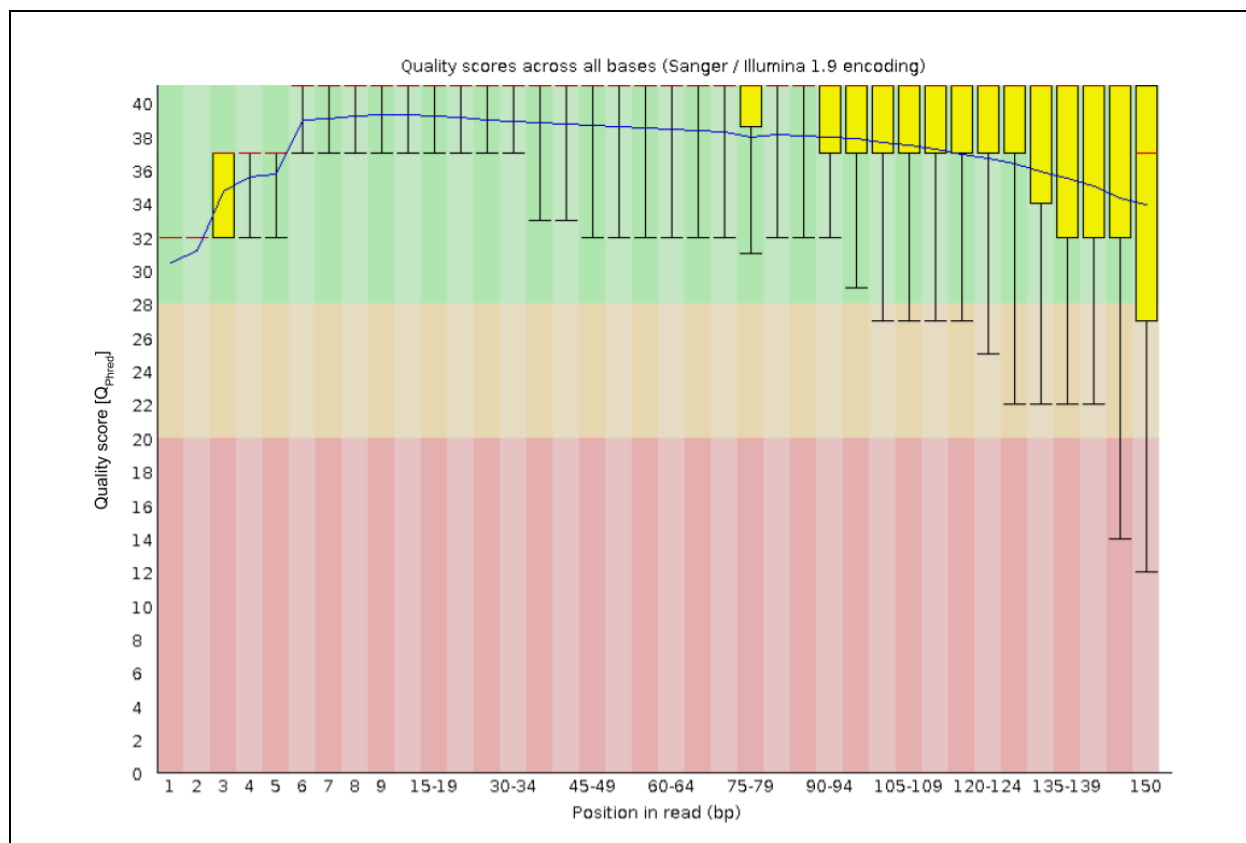


Figura S1-20: Puntajes promedio de calidad [Q_{phred}] por posición en la lectura para el exoma tumoral (lecturas *Reverse*).

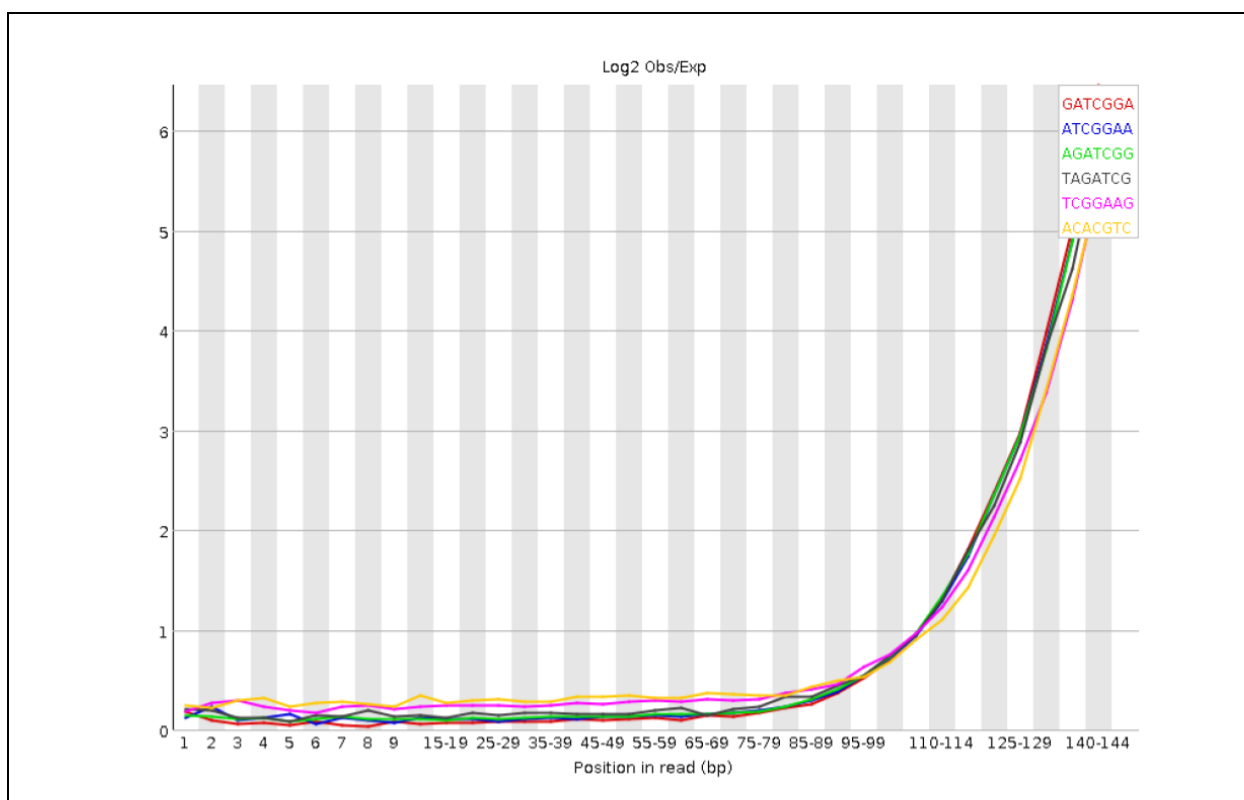


Figura S1-21: Distribución de *Kmers* por posición en la lectura para el exoma tumoral (lecturas *Forward*).

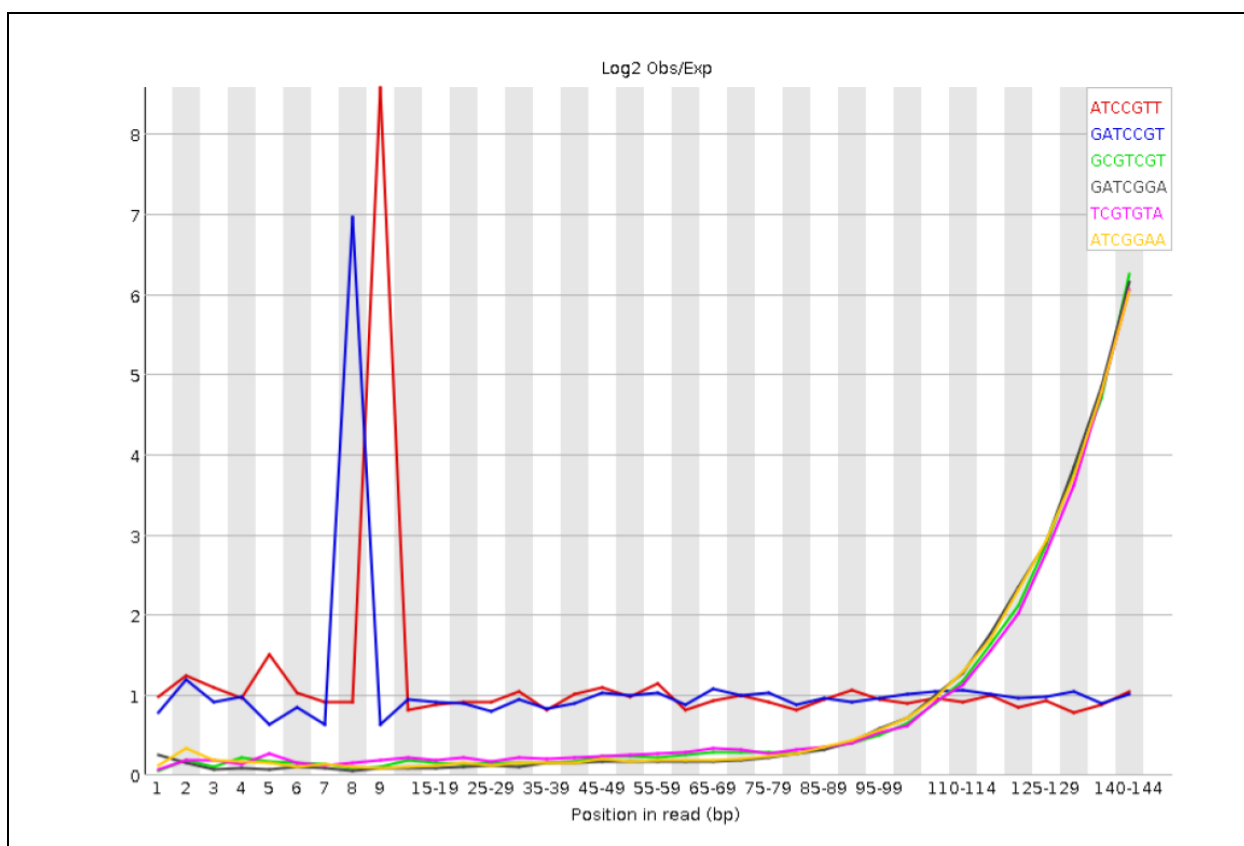


Figura S1-22: Distribución de *Kmers* por posición en la lectura para el exoma tumoral (lecturas *Reverse*).

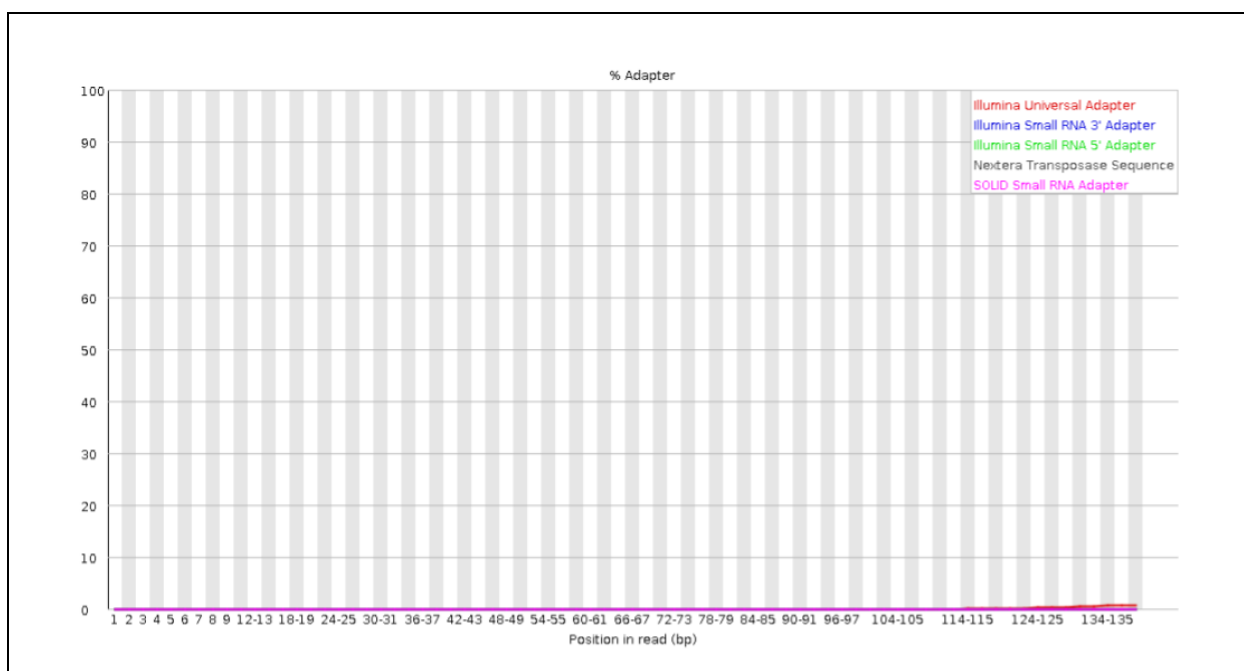


Figura S1-23: Distribución del contenido de adaptadores por posición en la lectura para el exoma tumoral (lecturas *Forward*).

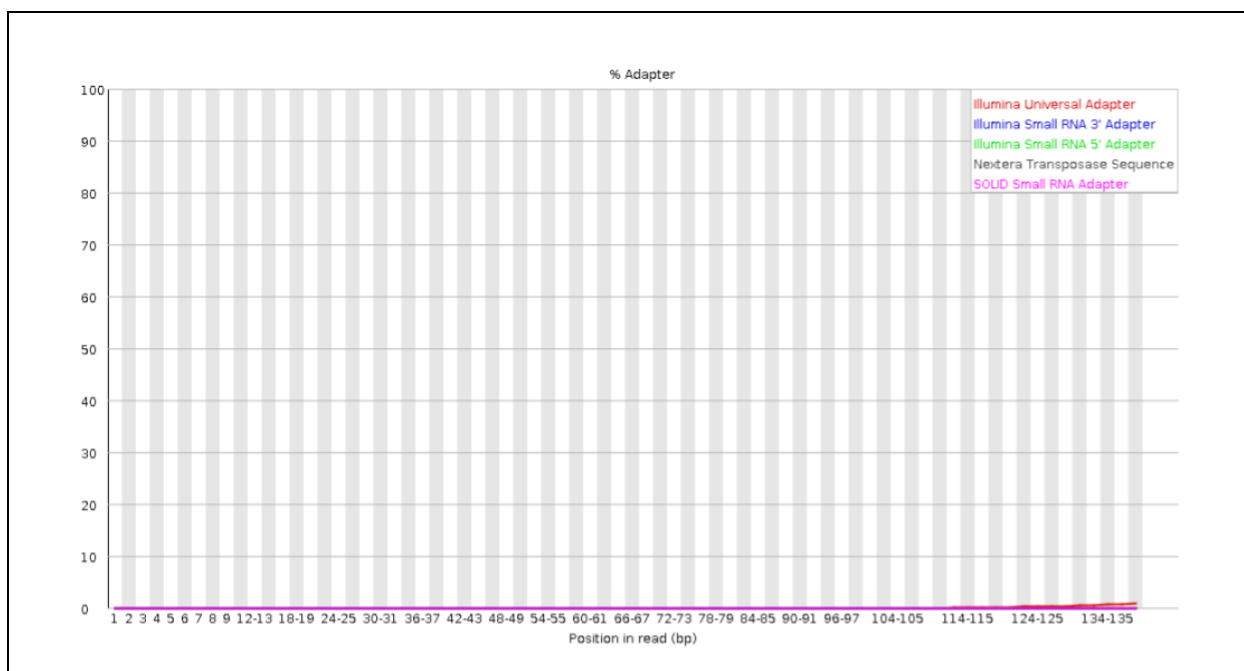


Figura S1-24: Distribución del contenido de adaptadores por posición en la lectura para el exoma tumoral (lecturas *Reverse*).

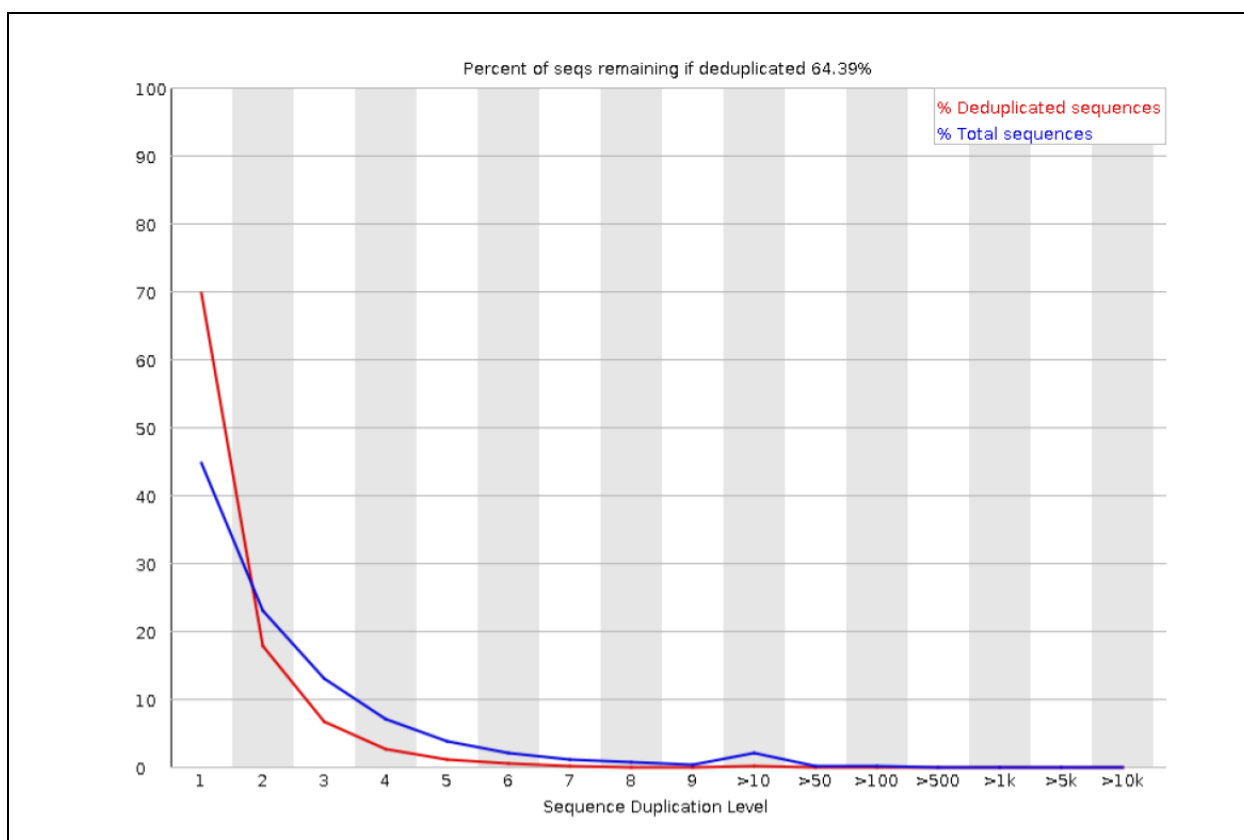


Figura S1-25: Distribución de los porcentajes de lecturas duplicadas de acuerdo al nivel de duplicación para el exoma tumoral (lecturas *Forward*).

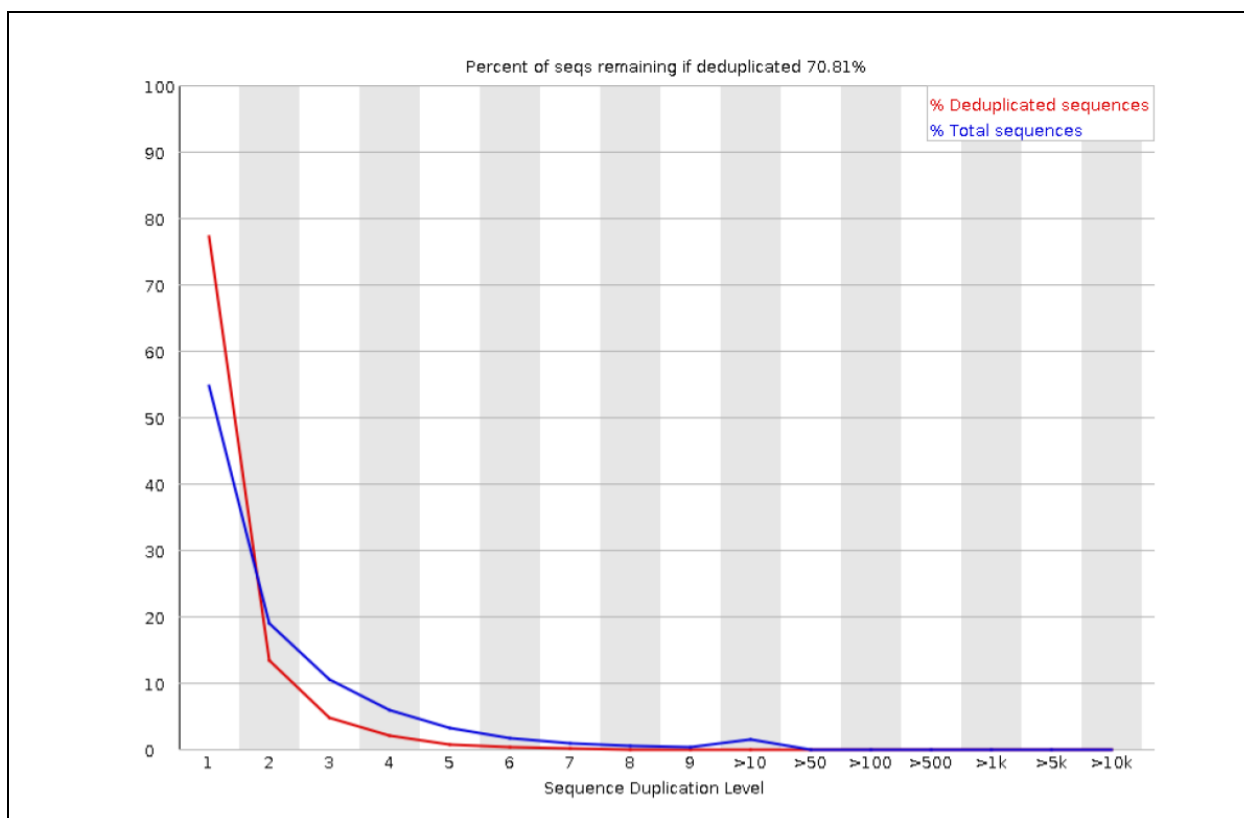


Figura S1-26: Distribución de los porcentajes de lecturas duplicadas de acuerdo al nivel de duplicación para el exoma tumoral (lecturas *Reverse*).

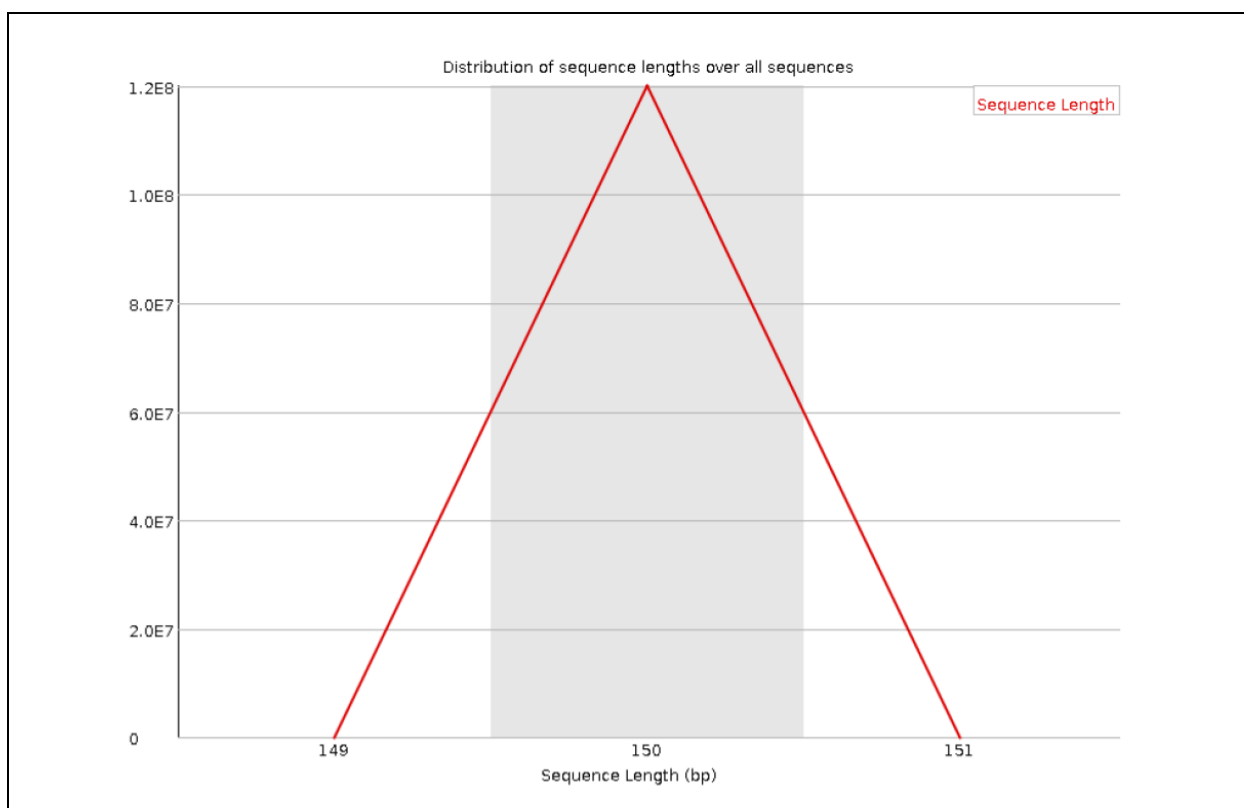


Figura S1-27: Distribución de las longitudes de lecturas para el exoma tumoral (lecturas *Forward*).

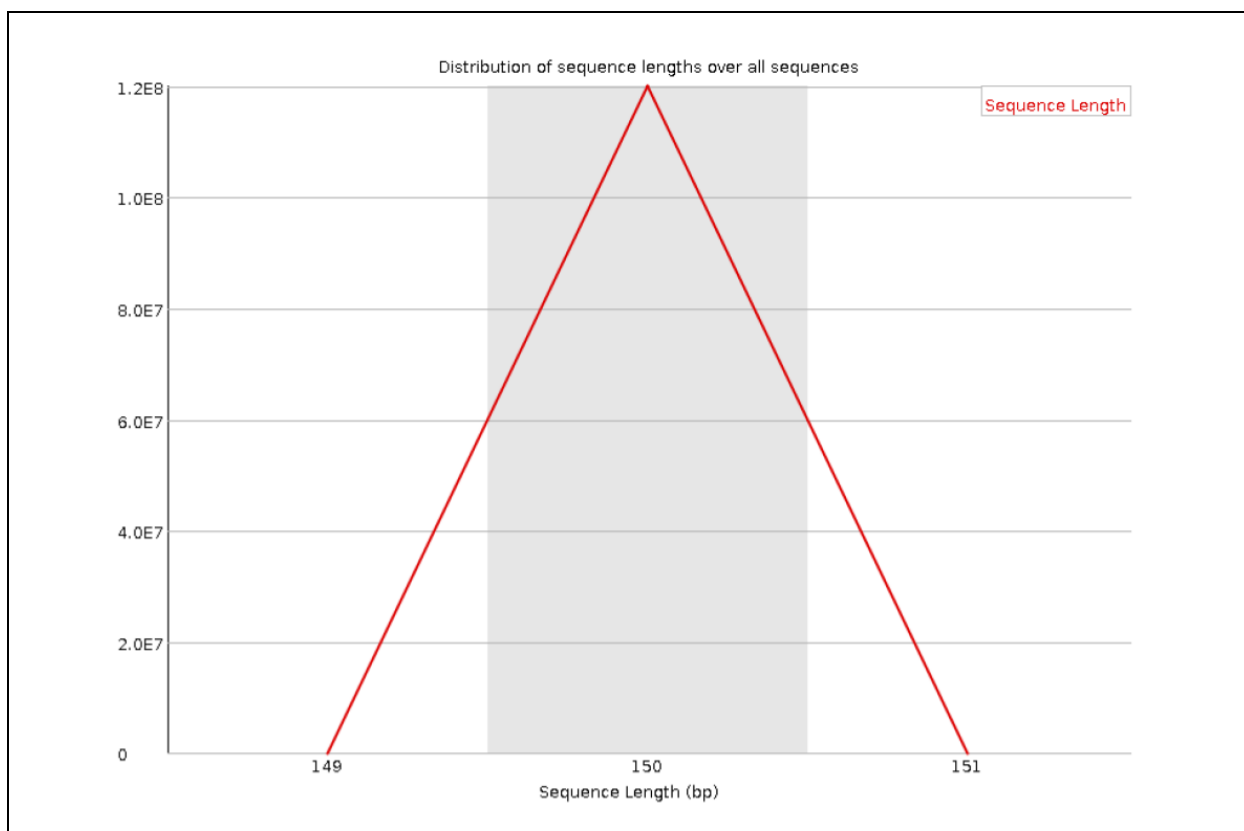


Figura S1-28: Distribución de las longitudes de lecturas para el exoma tumoral (lecturas *Reverse*).

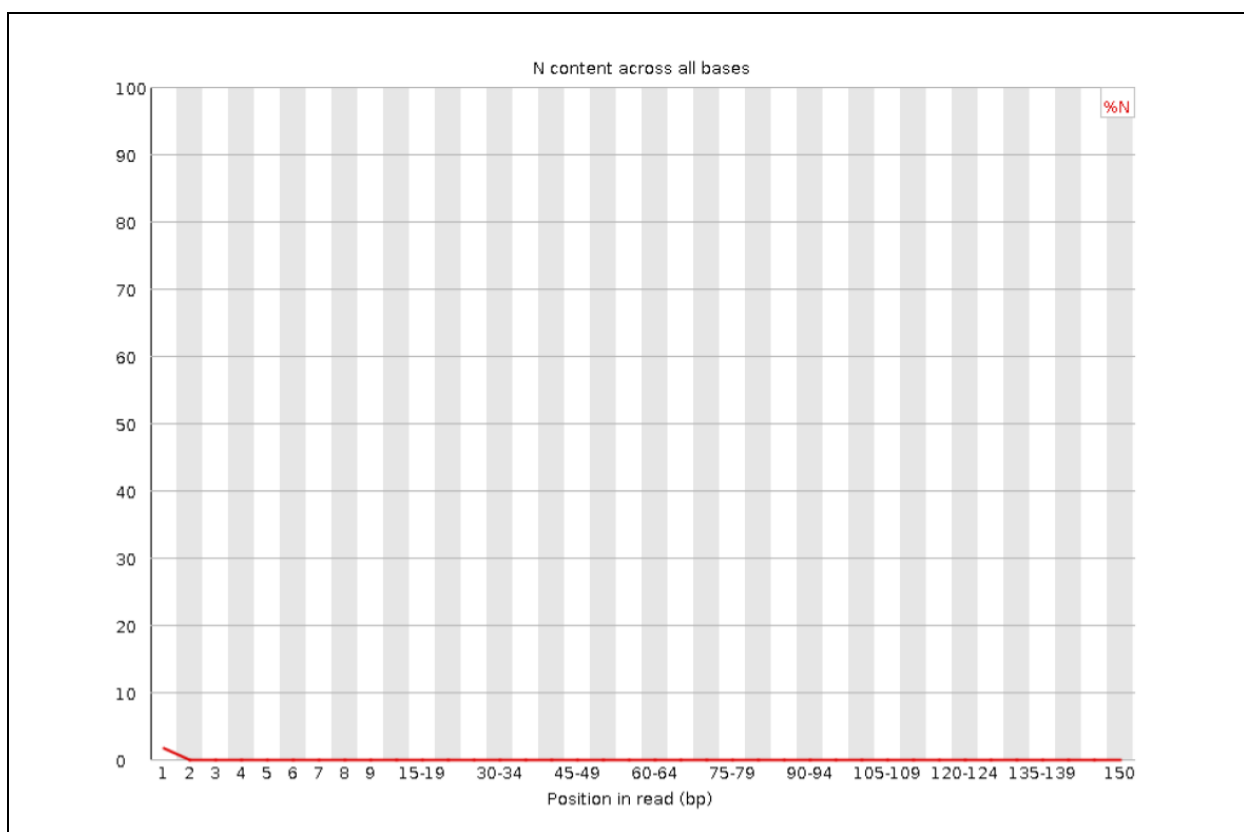


Figura S1-29: Distribución de los porcentajes de bases no identificadas por posición en la lectura para el exoma tumoral (lecturas *Forward*).

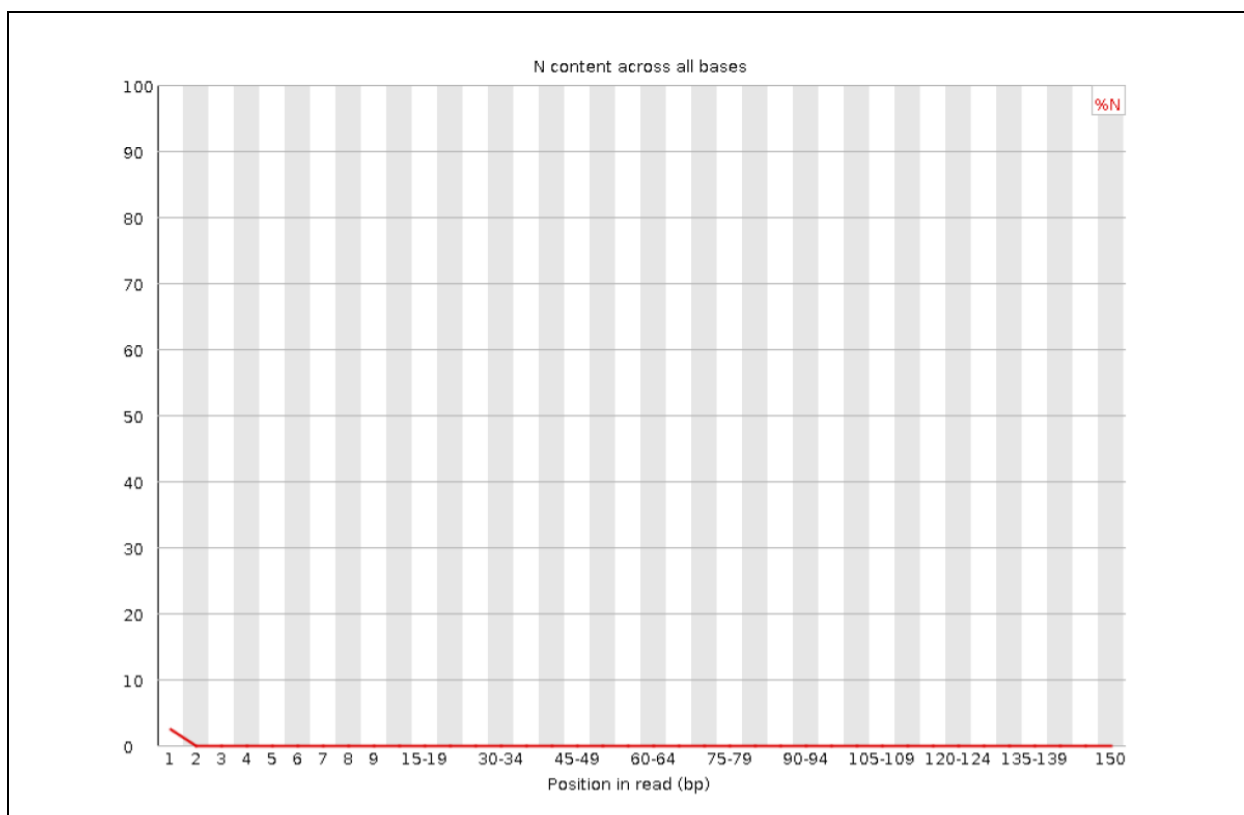


Figura S1-30: Distribución de los porcentajes de bases no identificadas por posición en la lectura para el exoma tumoral (lecturas *Reverse*).

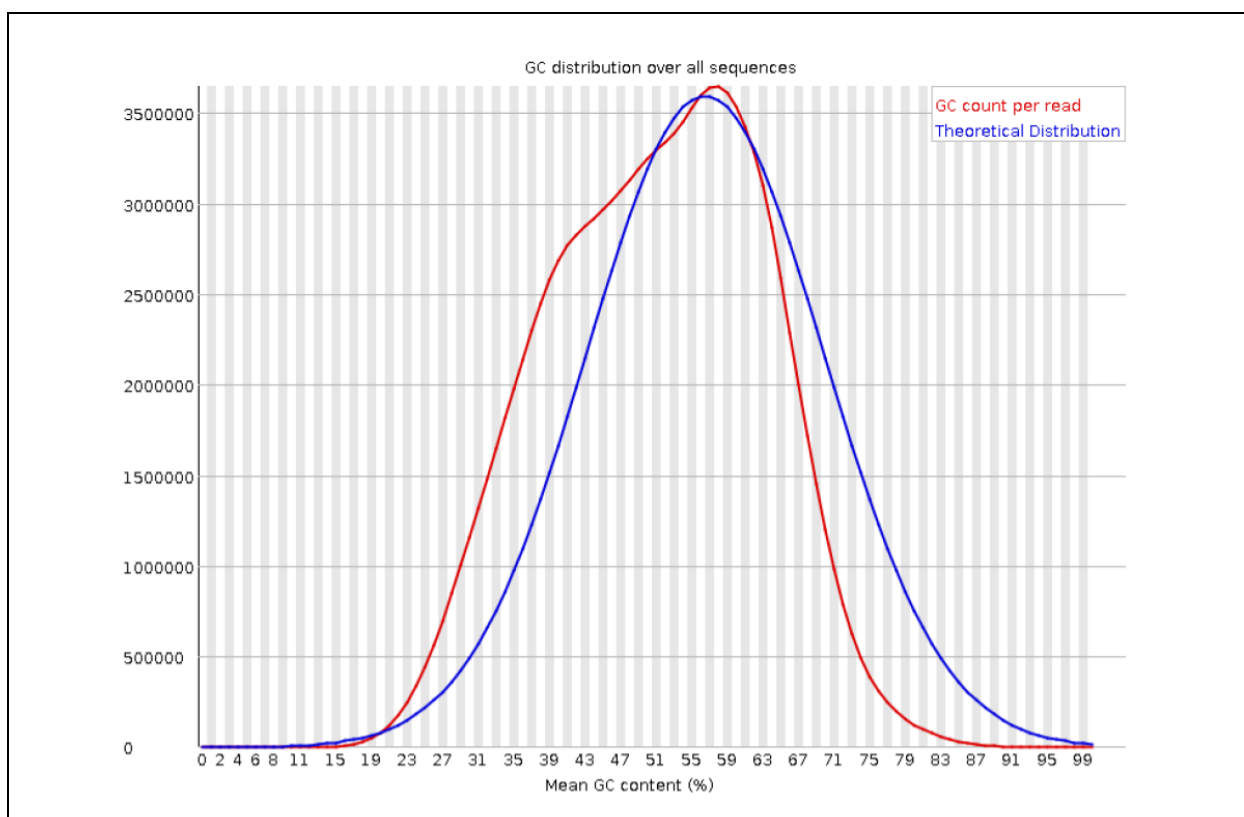


Figura S1-31: Distribución del contenido medio de GC en las lecturas para el exoma tumoral (lecturas *Forward*).

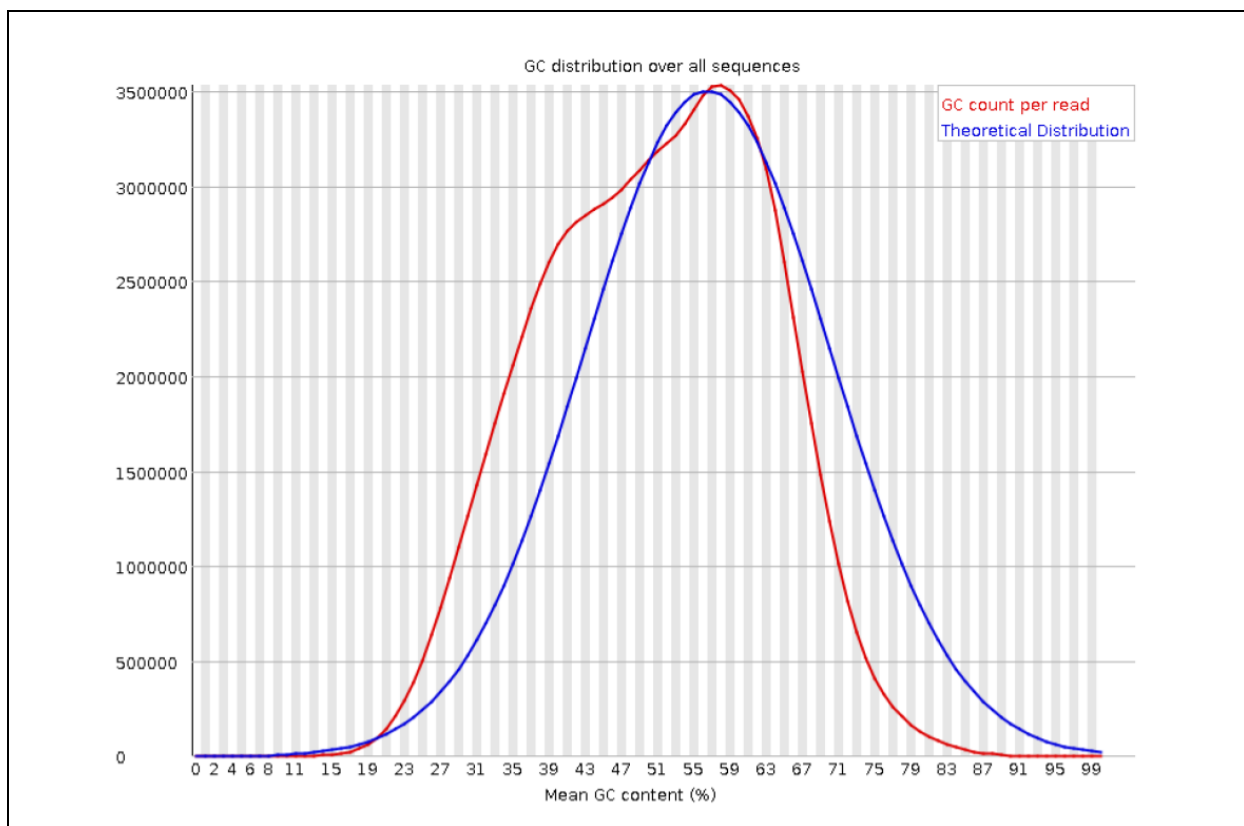


Figura S1-32: Distribución del contenido medio de GC en las lecturas para el exoma tumoral (lecturas *Reverse*).

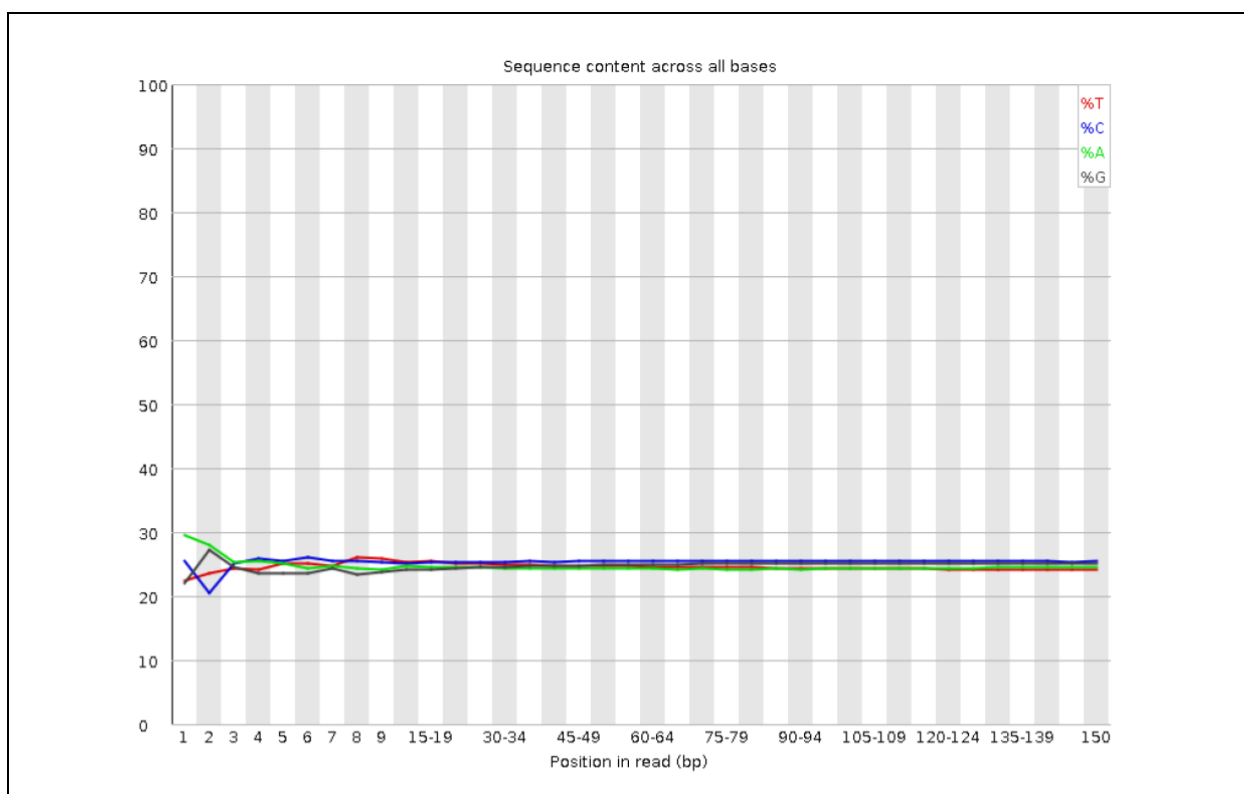


Figura S1-33: Distribución del contenido medio de GC por posición en la lectura para el exoma tumoral (lecturas *Forward*).

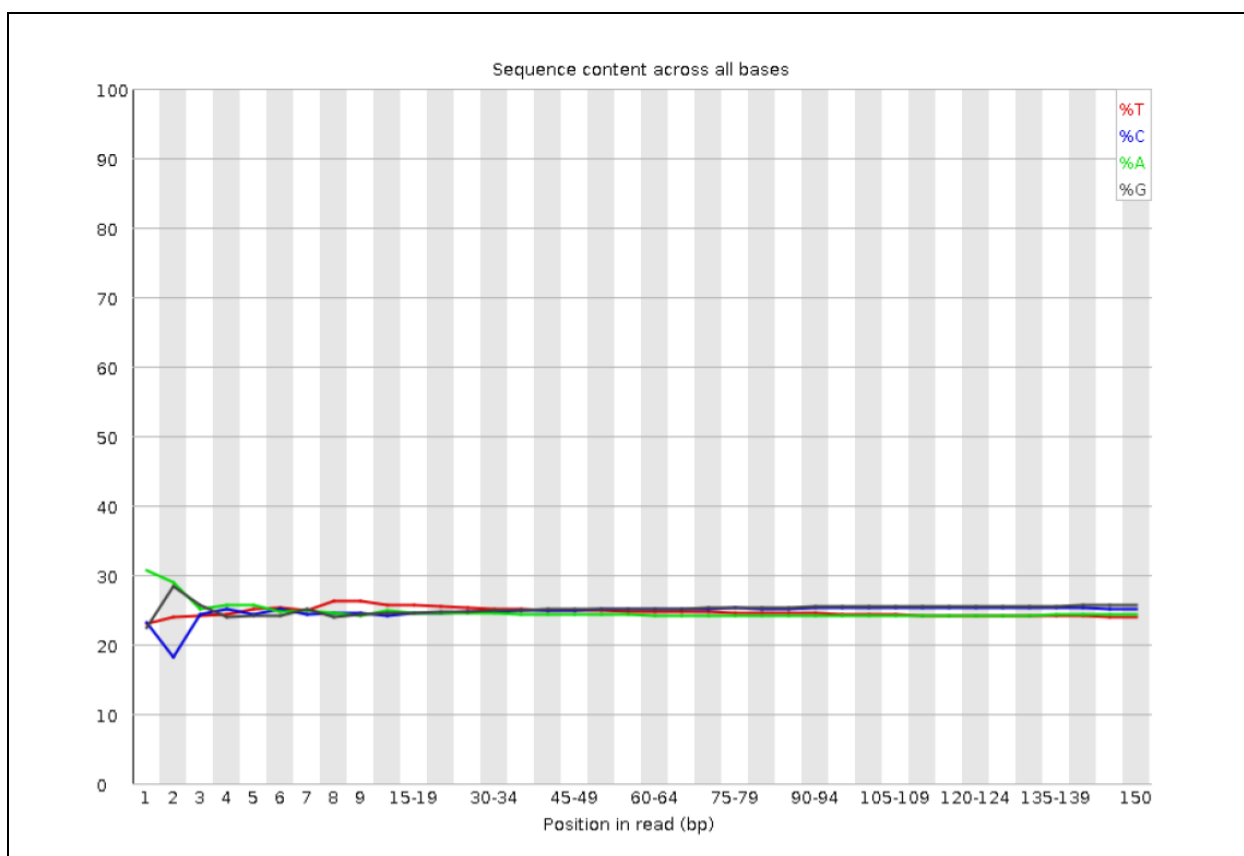


Figura S1-34: Distribución del contenido medio de GC por posición en la lectura para el exoma tumoral (lecturas *Reverse*).

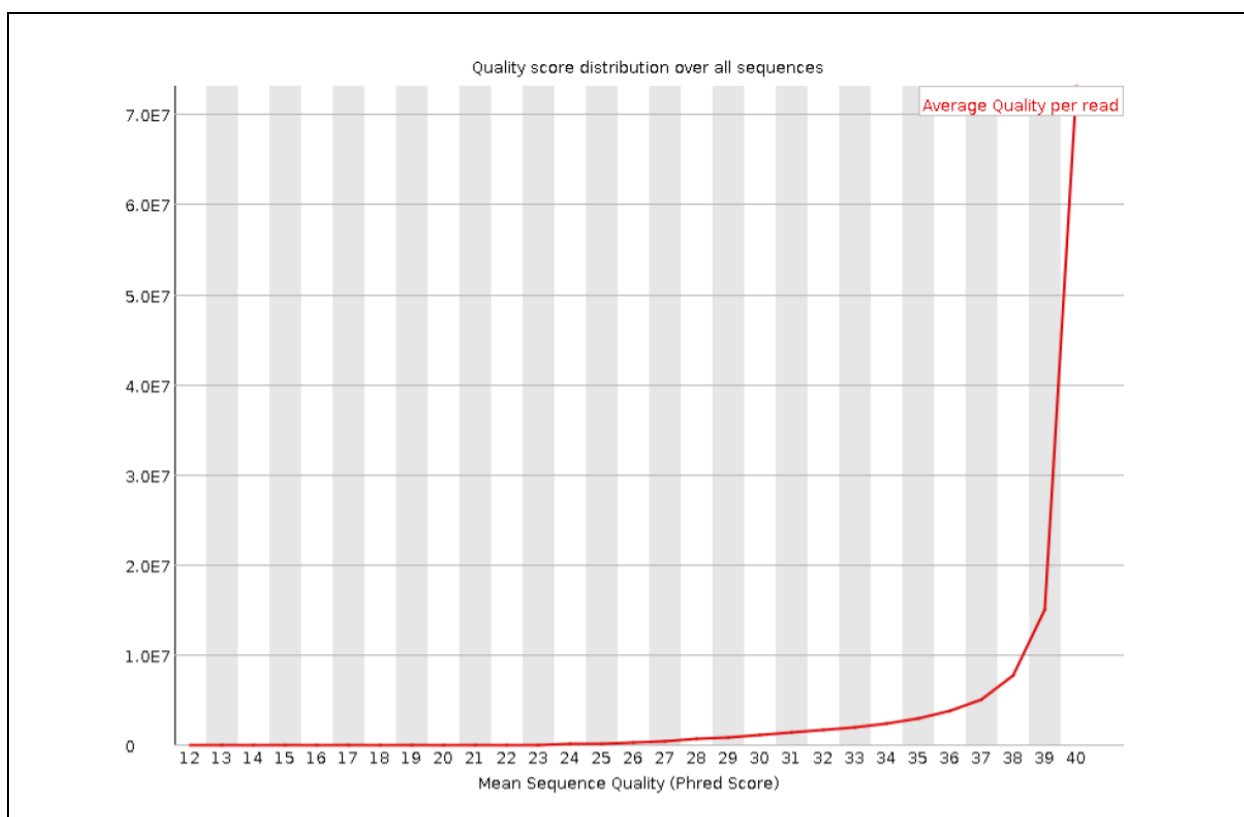


Figura S1-35: Distribución de los puntajes promedio de calidad por lectura para el exoma tumoral (lecturas *Forward*).

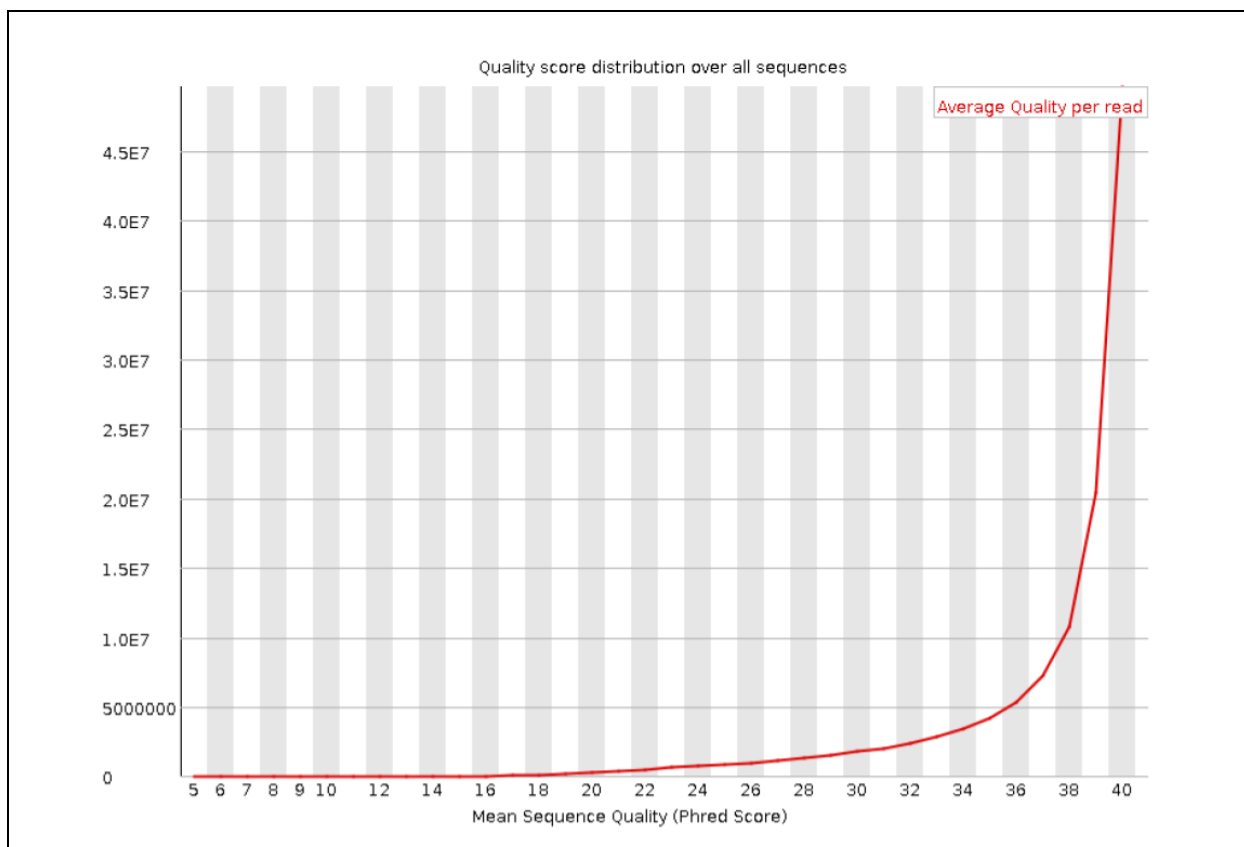


Figura S1-36: Distribución de los puntajes promedio de calidad por lectura para el exoma tumoral (lecturas *Reverse*).

Transcriptoma tumoral:

Estadísticas generales:

Lecturas <i>Forward</i>		Lecturas <i>Reverse</i>	
Magnitud	Valor	Magnitud	Valor
Lecturas totales	144'123.061	Lecturas totales	120'041.101
Lecturas de baja calidad	0	Lecturas de baja calidad	0
Longitud lectura	150	Longitud lectura	150
% GC	53	% GC	53

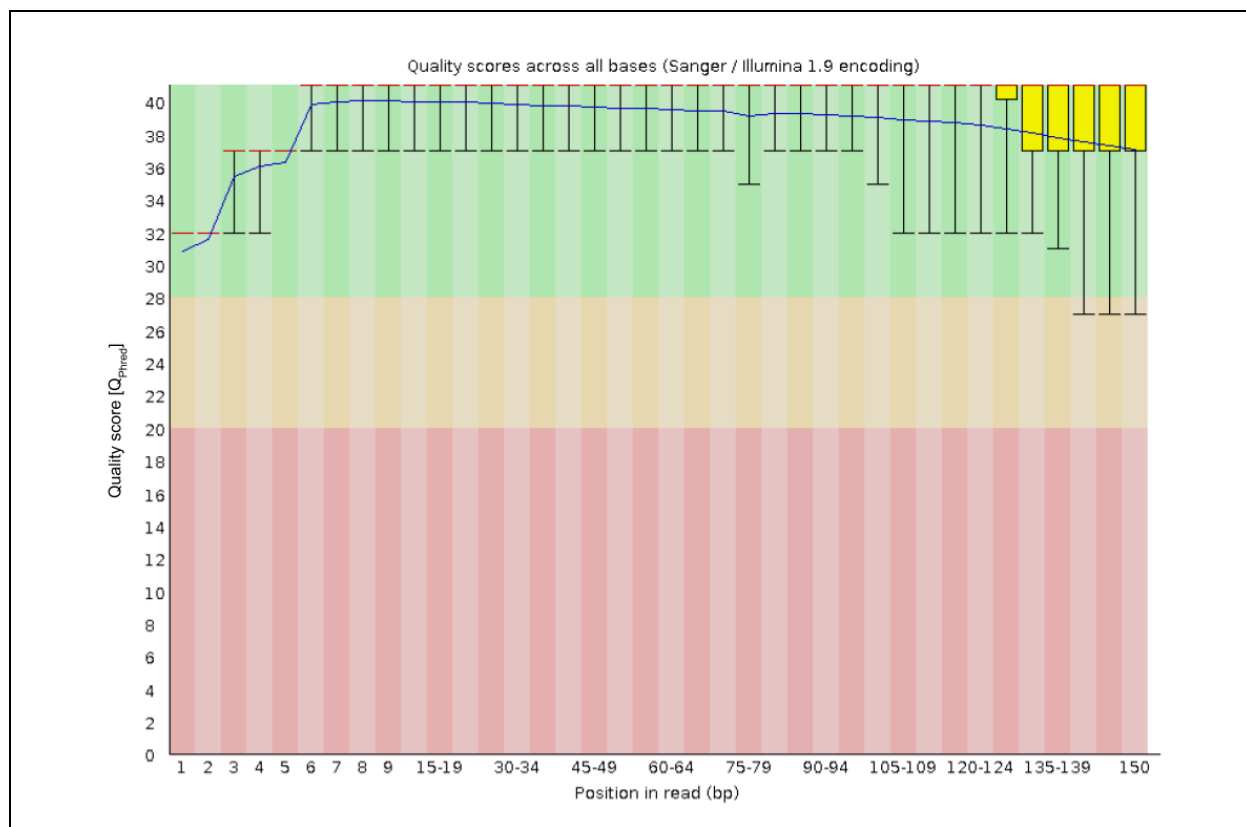


Figura S1-37: Puntajes promedio de calidad [Q_{phred}] por posición en la lectura para el transcriptoma tumoral (lecturas *Forward*).

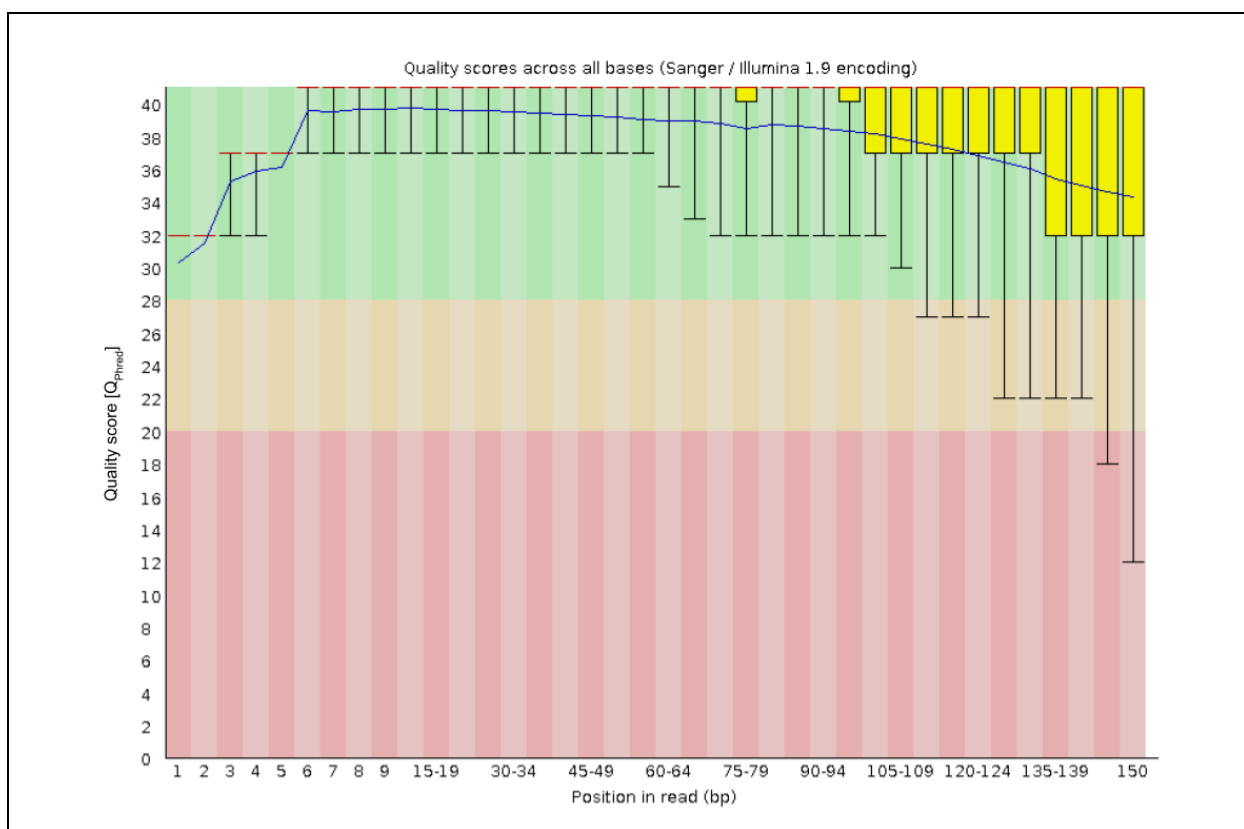


Figura S1-38: Puntajes promedio de calidad [Q_{phred}] por posición en la lectura para el transcriptoma tumoral (lecturas *Reverse*).

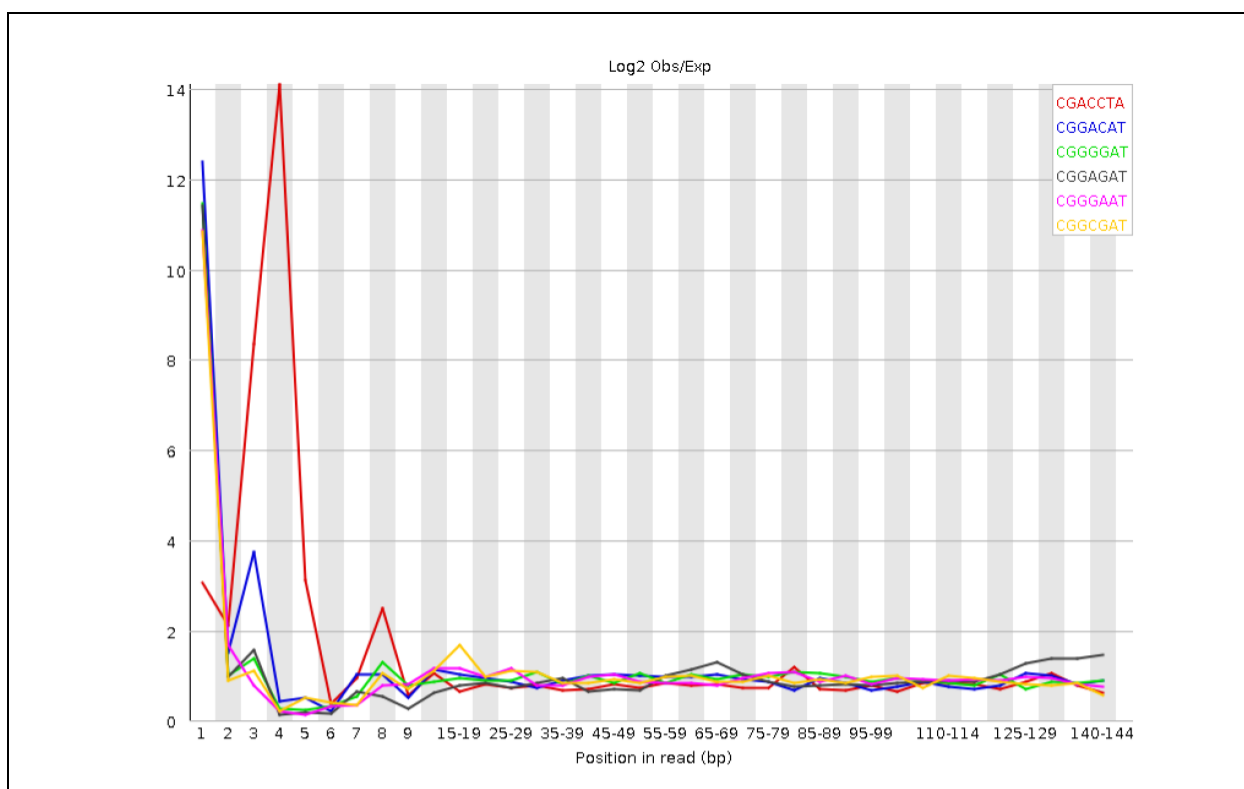


Figura S1-39: Distribución de *Kmers* por posición en la lectura para el transcriptoma tumoral (lecturas *Forward*).

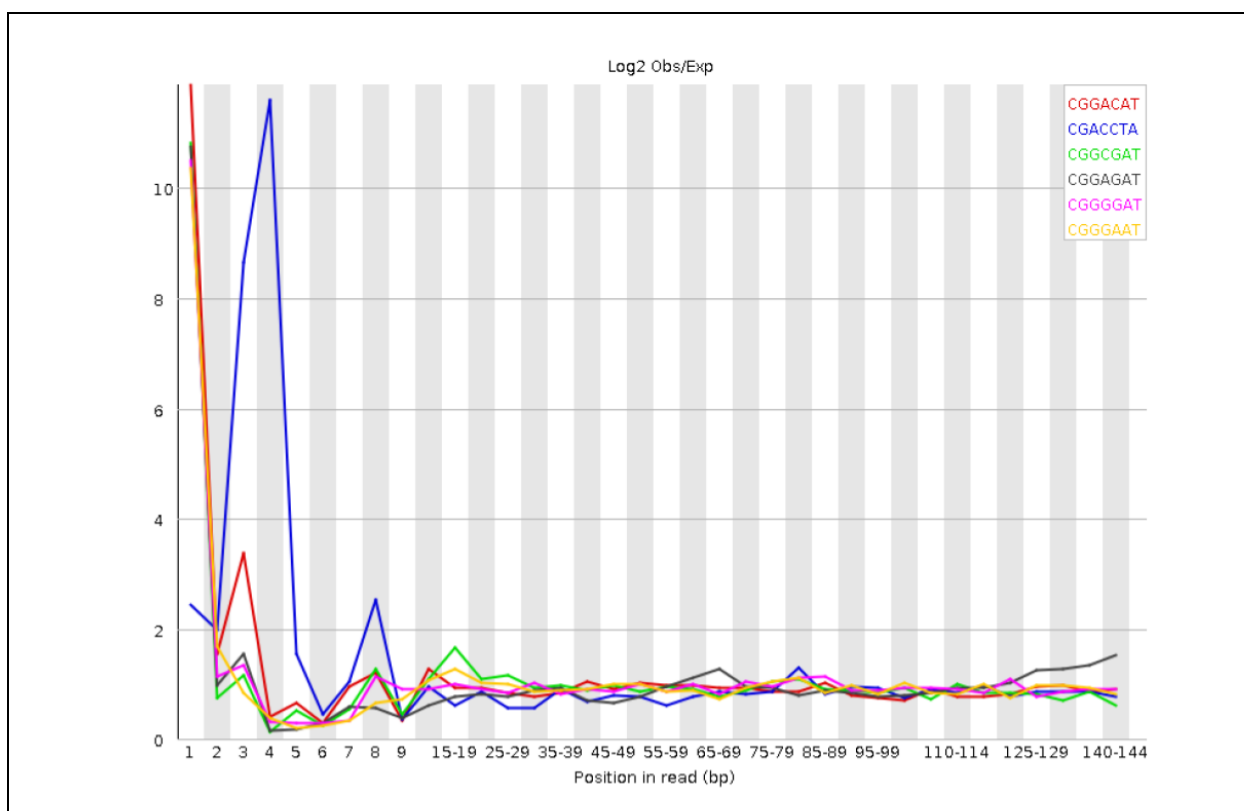


Figura S1-40: Distribución de *Kmers* por posición en la lectura para el transcriptoma tumoral (lecturas *Reverse*).

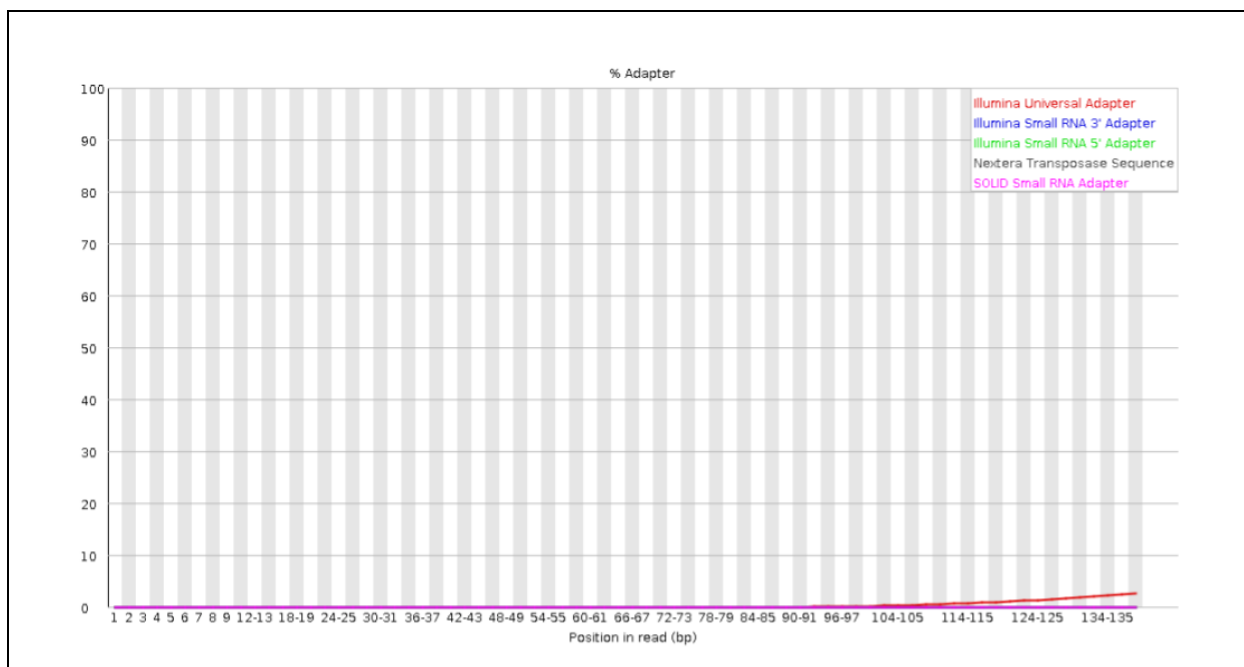


Figura S1-41: Distribución del contenido de adaptadores por posición en la lectura para el transcriptoma tumoral (lecturas *Forward*).

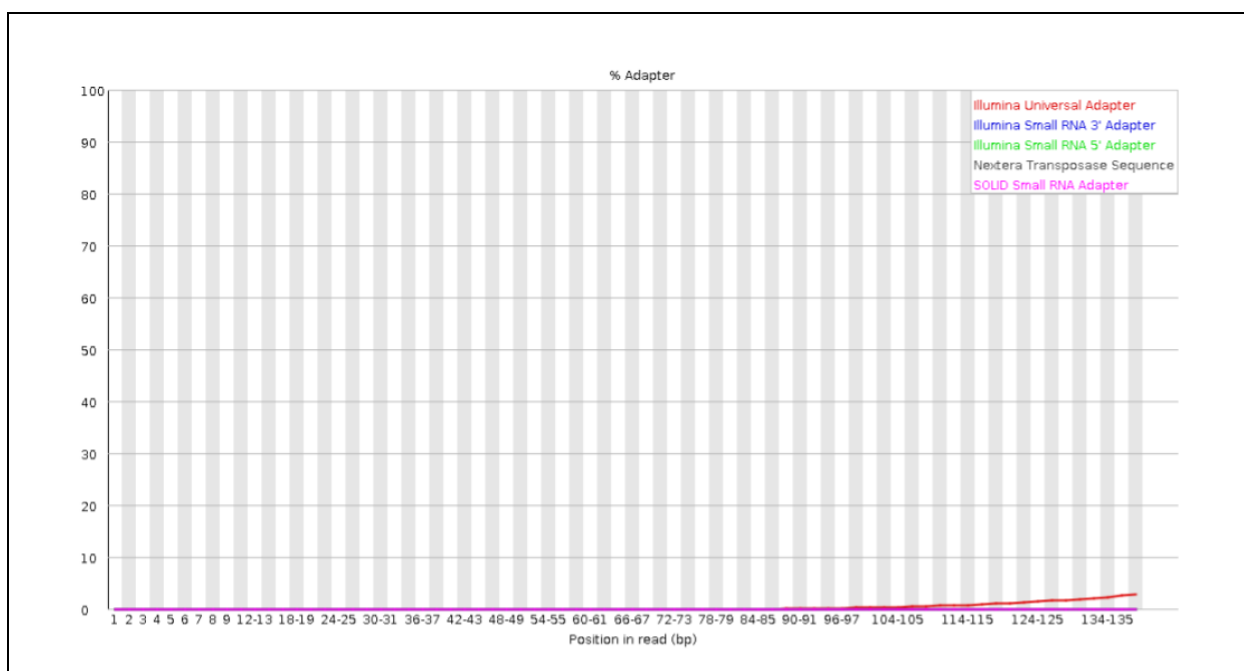


Figura S1-42: Distribución del contenido de adaptadores por posición en la lectura para el transcriptoma tumoral (lecturas *Reverse*).

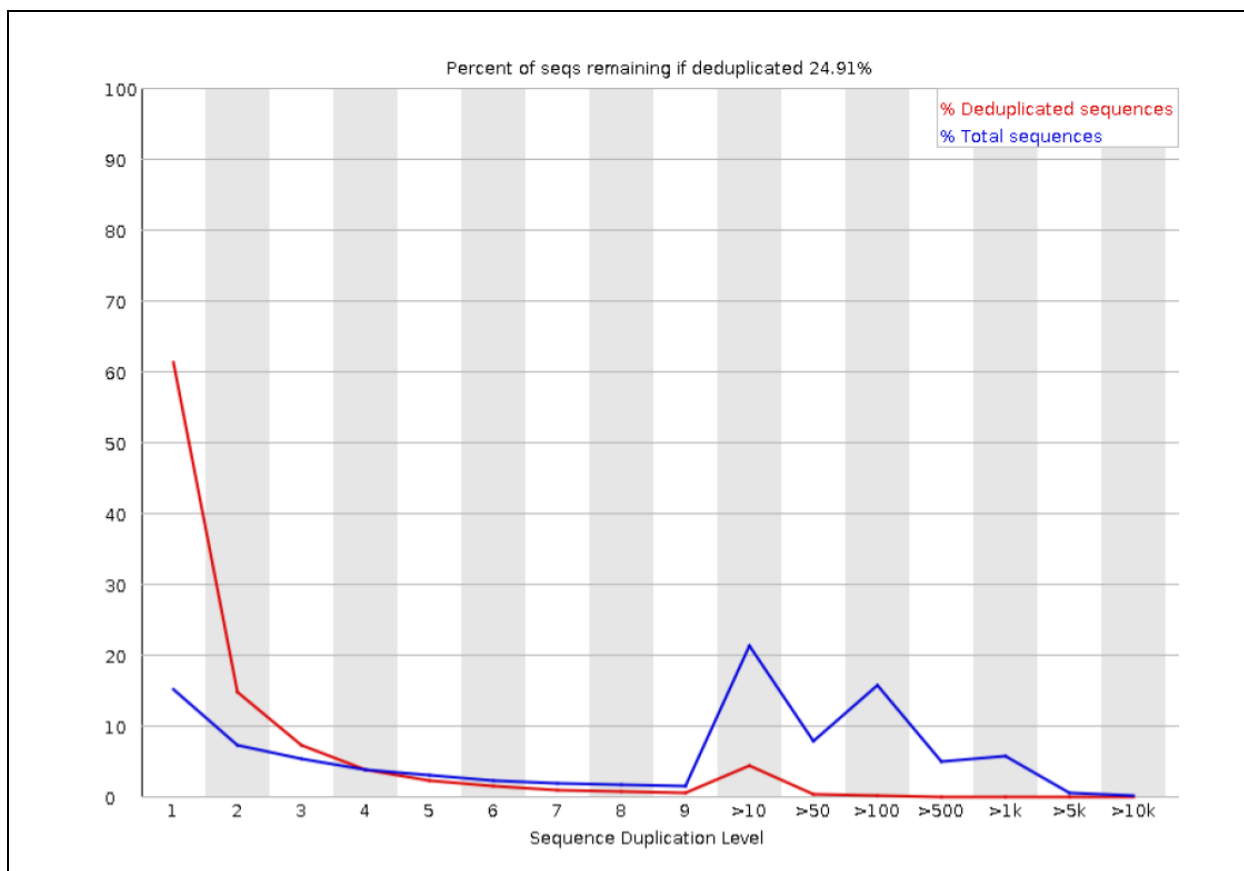


Figura S1-43: Distribución de los porcentajes de lecturas duplicadas de acuerdo al nivel de duplicación para el transcriptoma tumoral (lecturas *Forward*).

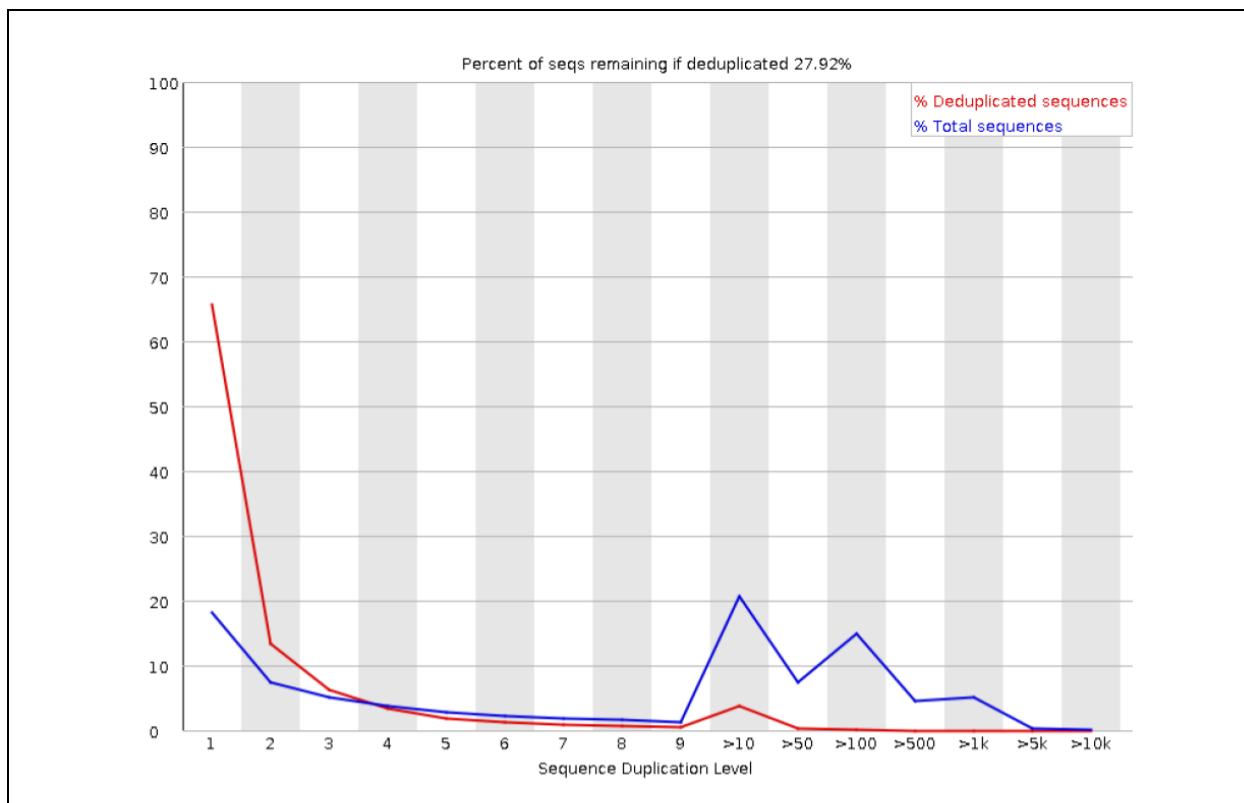


Figura S1-44: Distribución de los porcentajes de lecturas duplicadas de acuerdo al nivel de duplicación para el transcriptoma tumoral (lecturas *Reverse*).

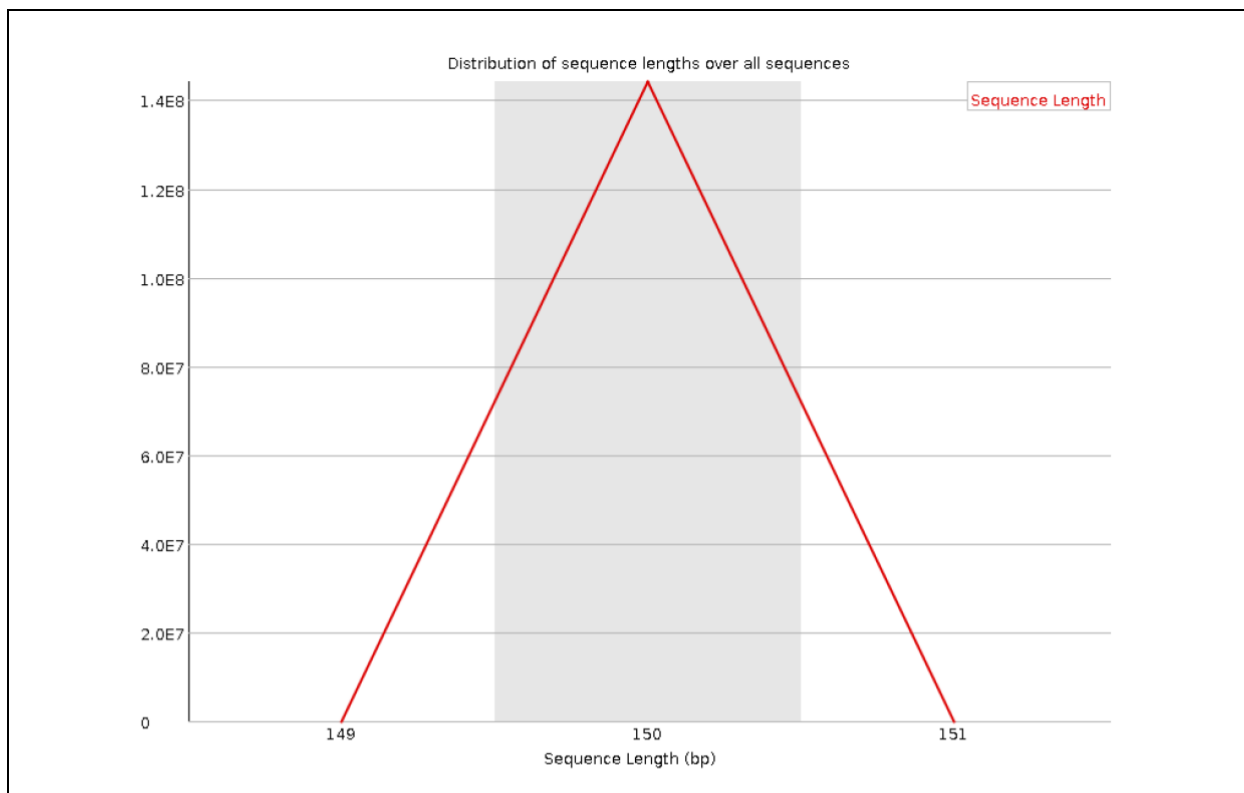


Figura S1-45: Distribución de las longitudes de lecturas para el transcriptoma tumoral (lecturas *Forward*).

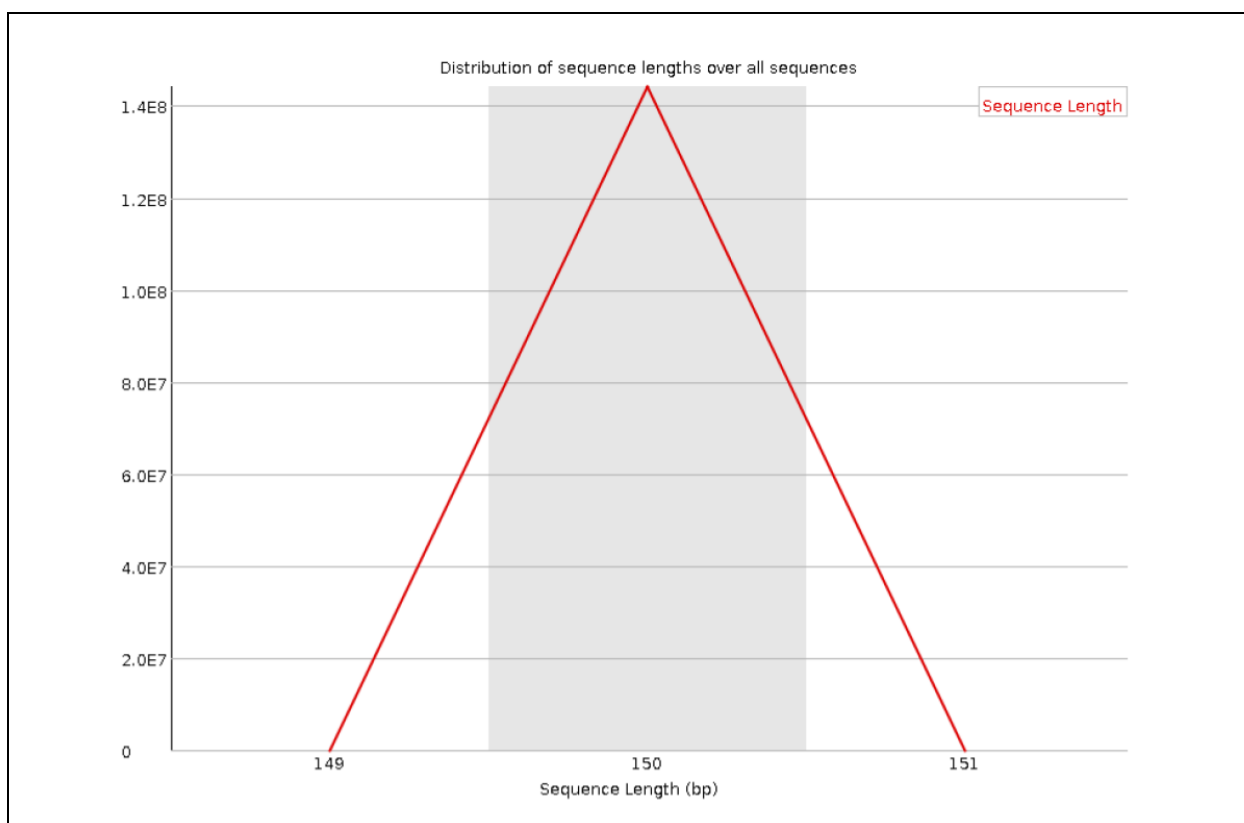


Figura S1-46: Distribución de las longitudes de lecturas para el exoma tumoral (lecturas *Reverse*).

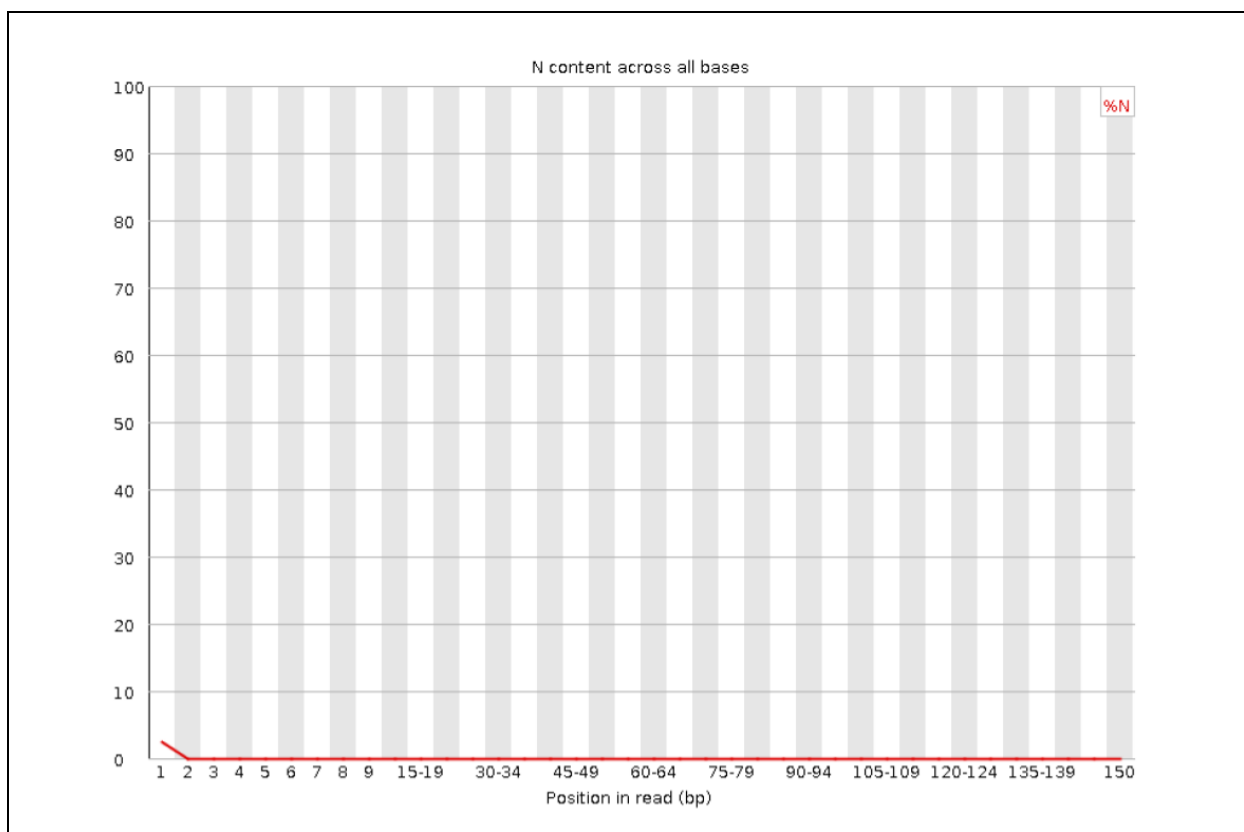


Figura S1-47: Distribución de los porcentajes de bases no identificadas por posición en la lectura para el transcriptoma tumoral (lecturas *Forward*).

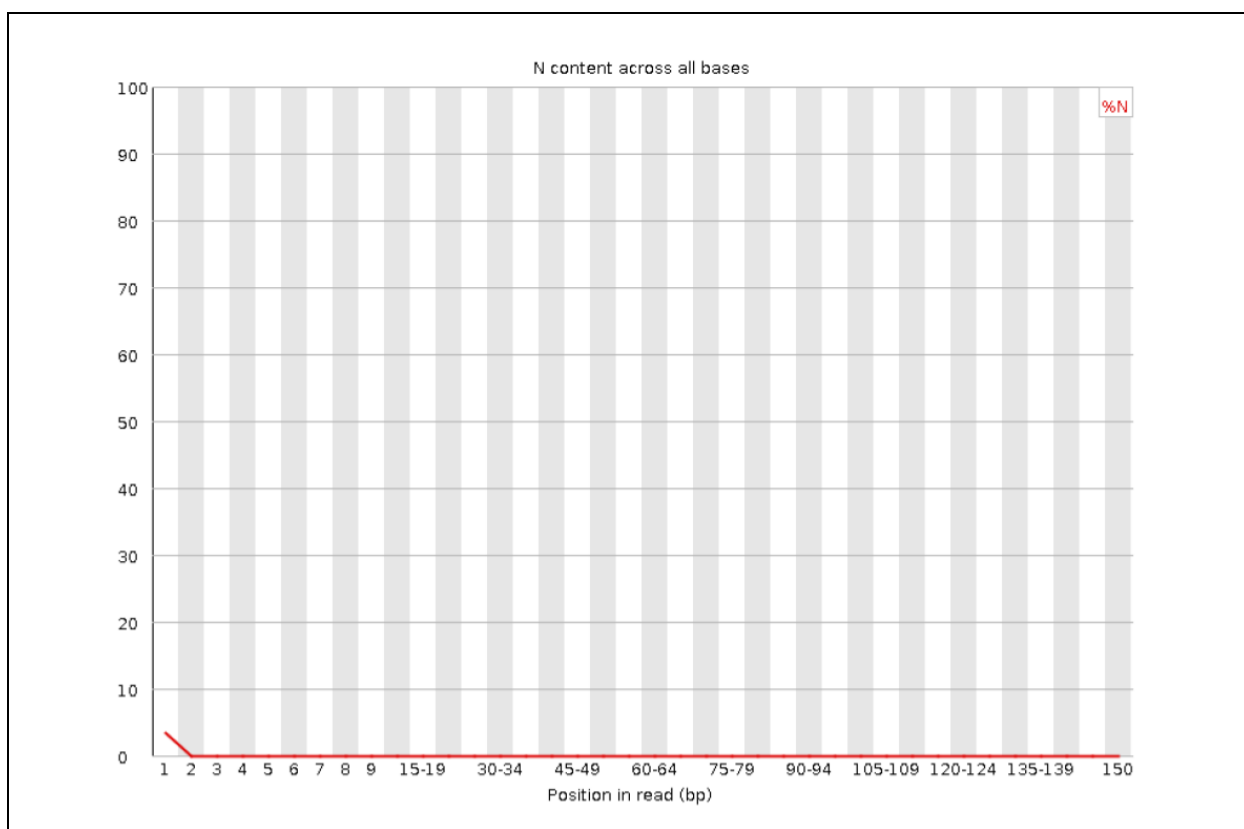


Figura S1-48: Distribución de los porcentajes de bases no identificadas por posición en la lectura para el transcriptoma tumoral (lecturas *Reverse*).

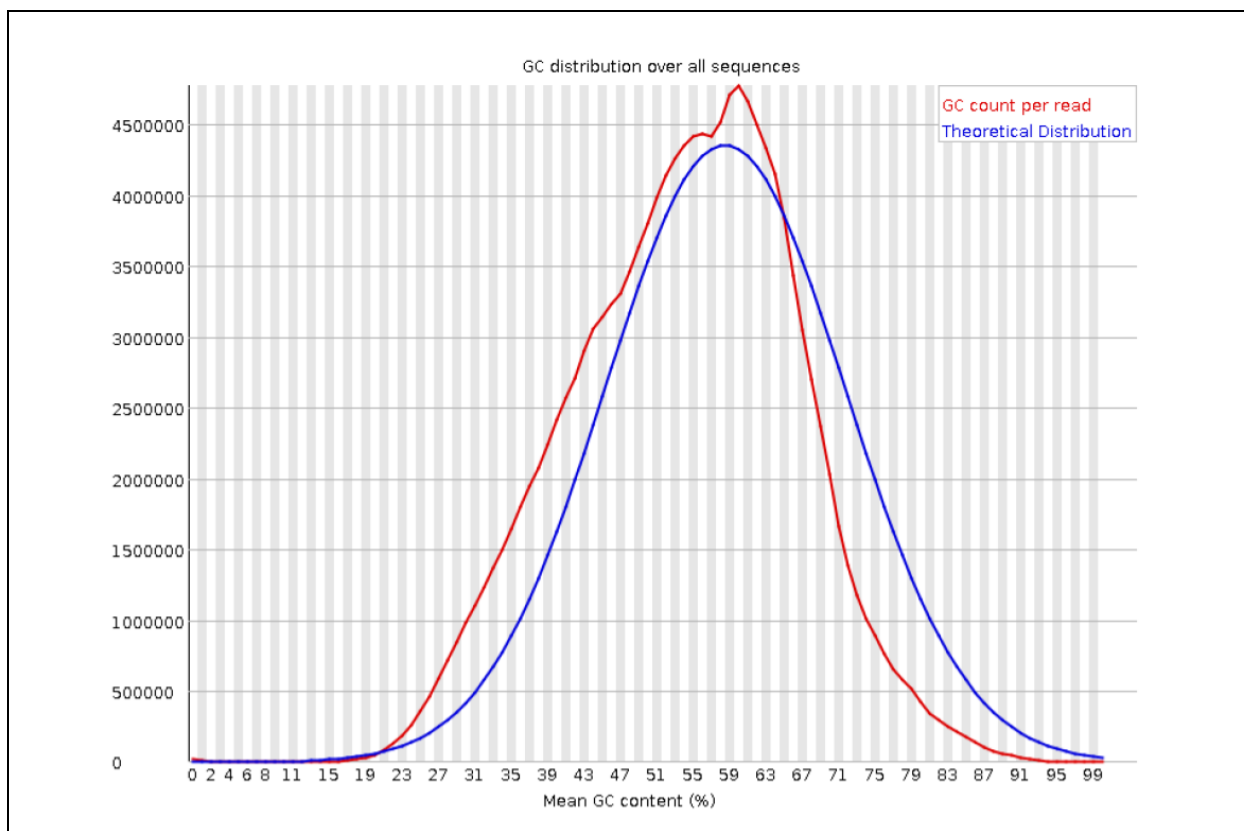


Figura S1-49: Distribución del contenido medio de GC en las lecturas para el transcriptoma tumoral (lecturas *Forward*).

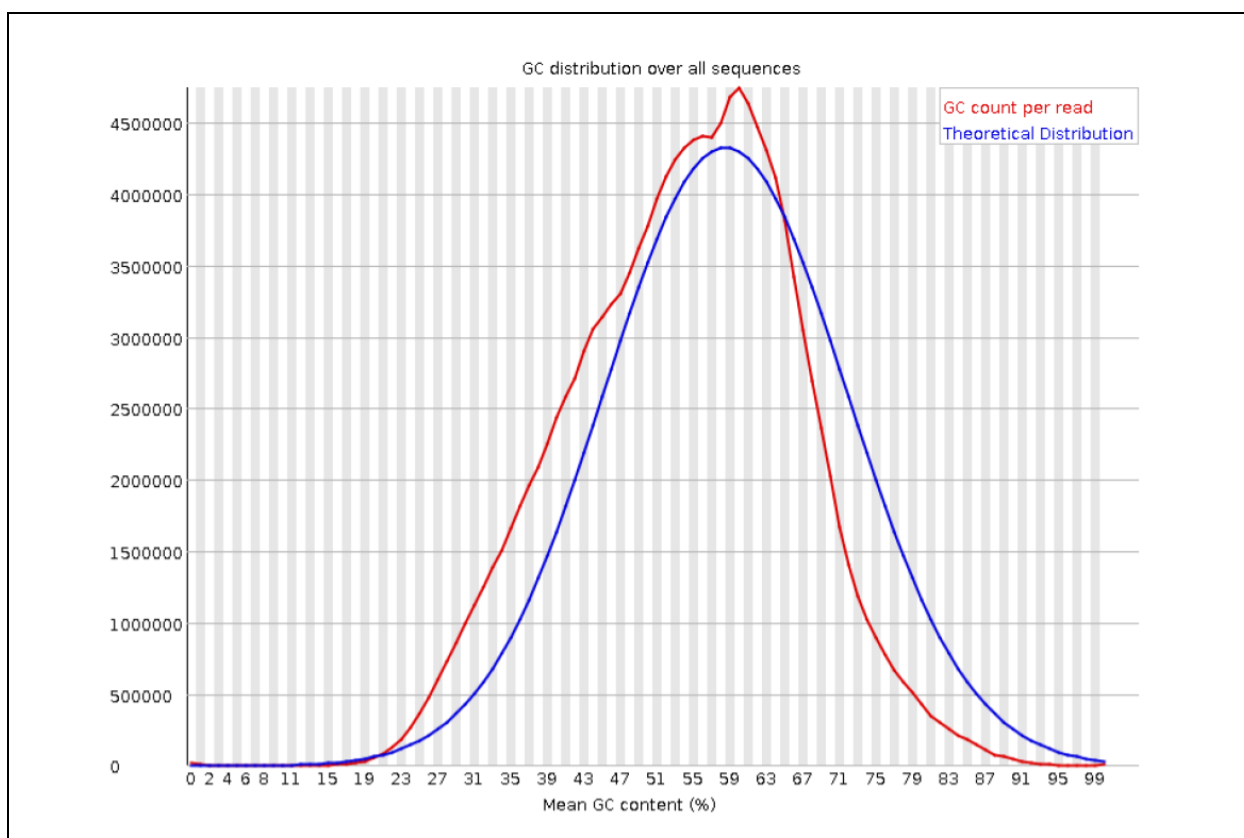


Figura S1-50: Distribución del contenido medio de GC en las lecturas para el transcriptoma tumoral (lecturas *Reverse*).

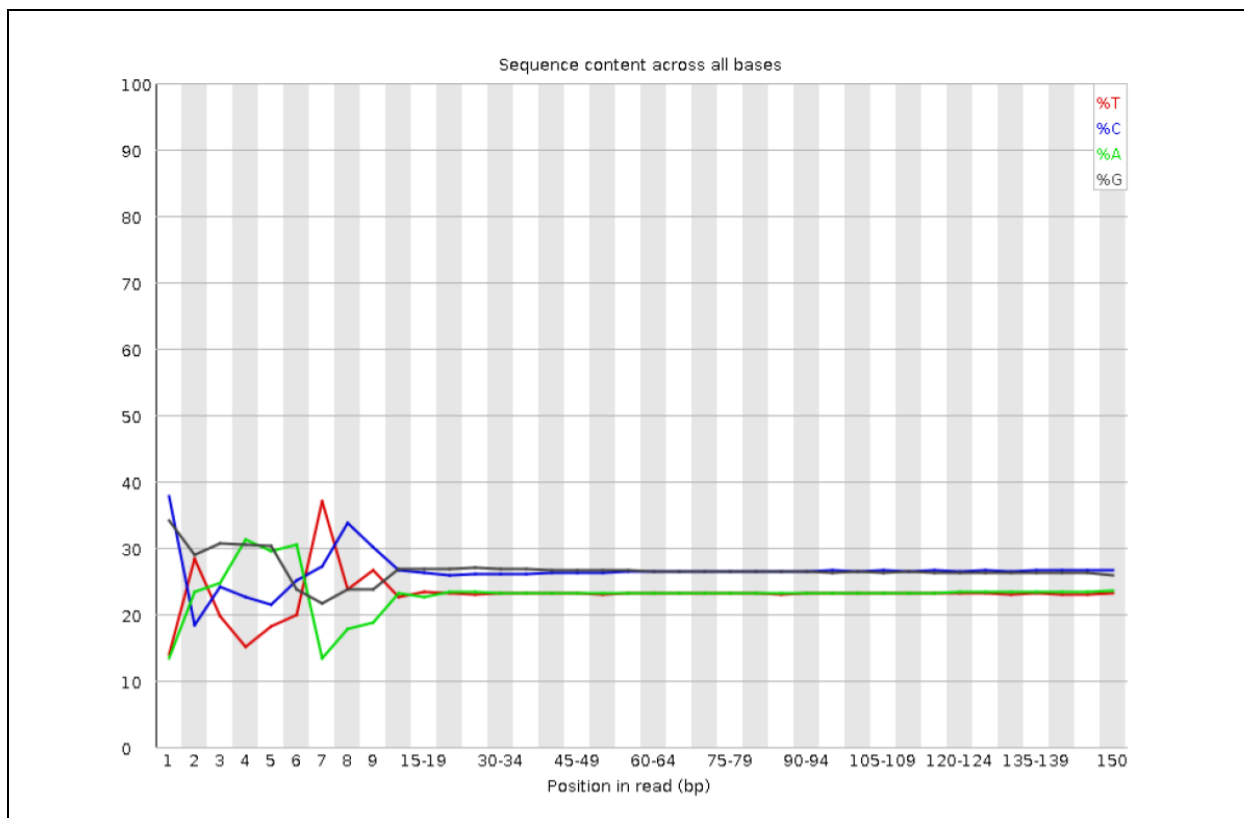


Figura S1-51: Distribución del contenido medio de GC por posición en la lectura para el transcriptoma tumoral (lecturas *Forward*).

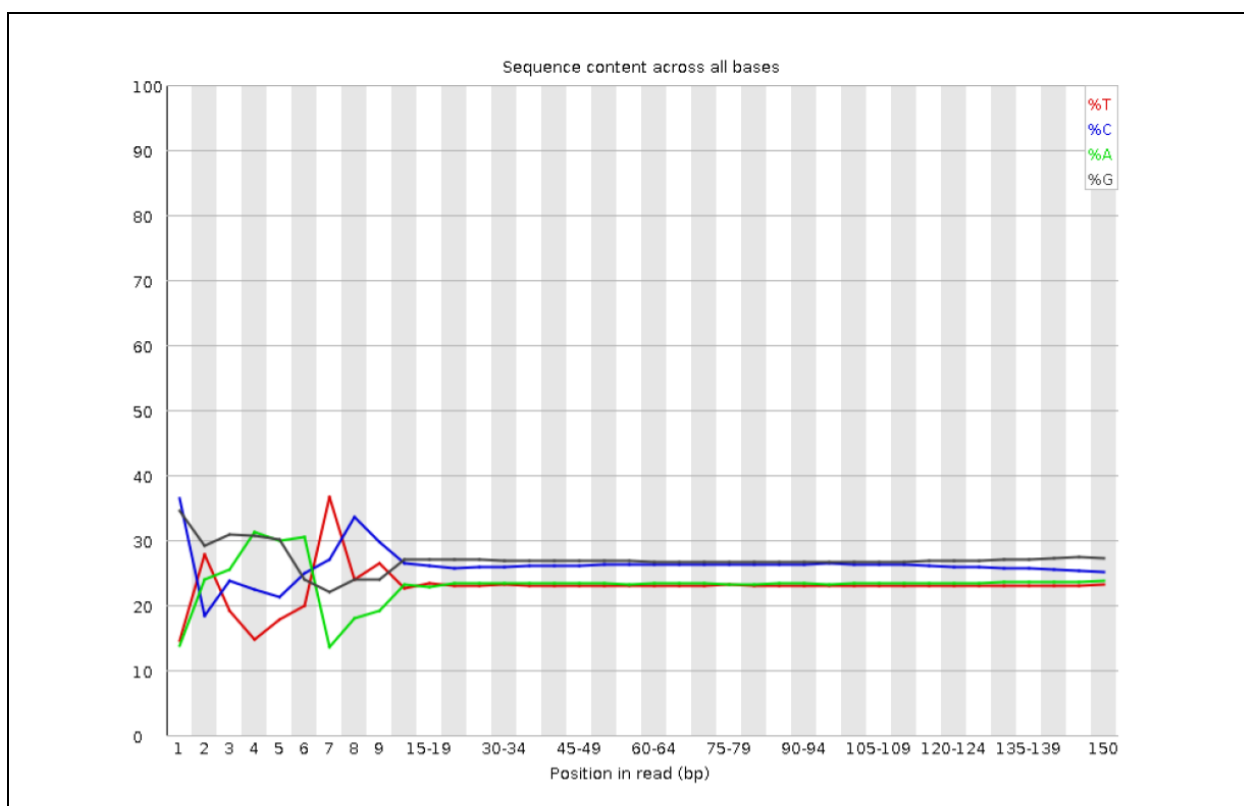


Figura S1-52: Distribución del contenido medio de GC por posición en la lectura para el transcriptoma tumoral (lecturas *Reverse*).

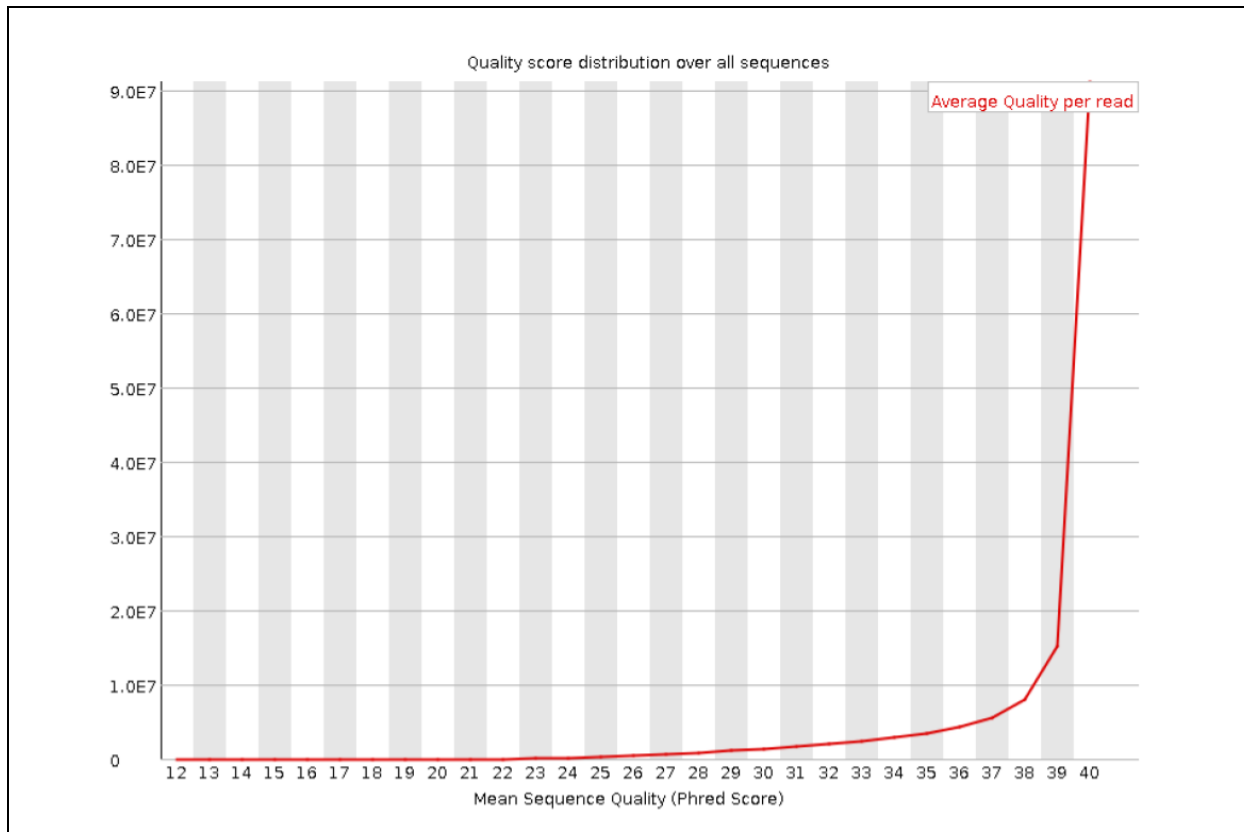


Figura S1-53: Distribución de los puntajes promedio de calidad por lectura para el transcriptoma tumoral (lecturas *Forward*).

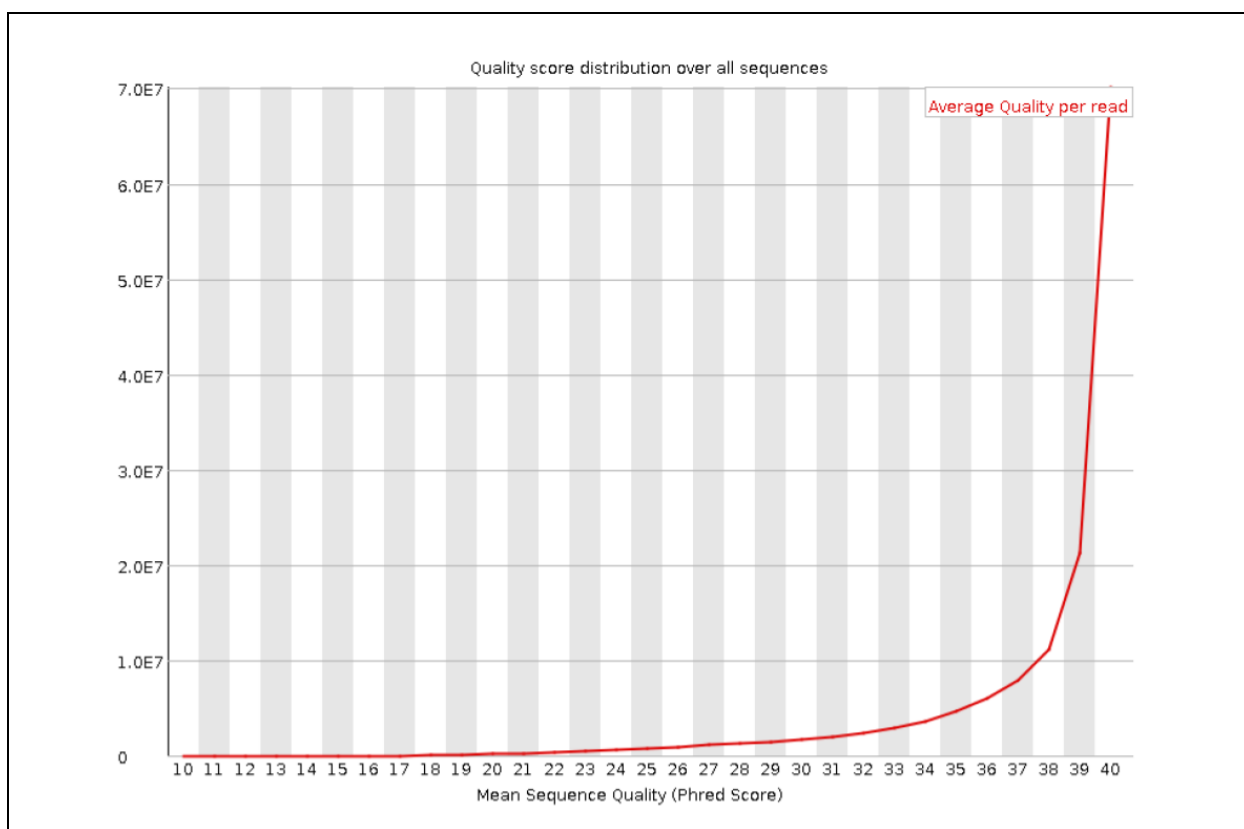


Figura S1-54: Distribución de los puntajes promedio de calidad por lectura para el transcriptoma tumoral (lecturas *Reverse*).

Anexo 2: Resultados tipificación haplotipo HLA

Resultados tipificación con la herramienta optiType:

En la **Figura S2-1** se presentan los resultados de los alineamientos obtenidos para los genes tipificados.

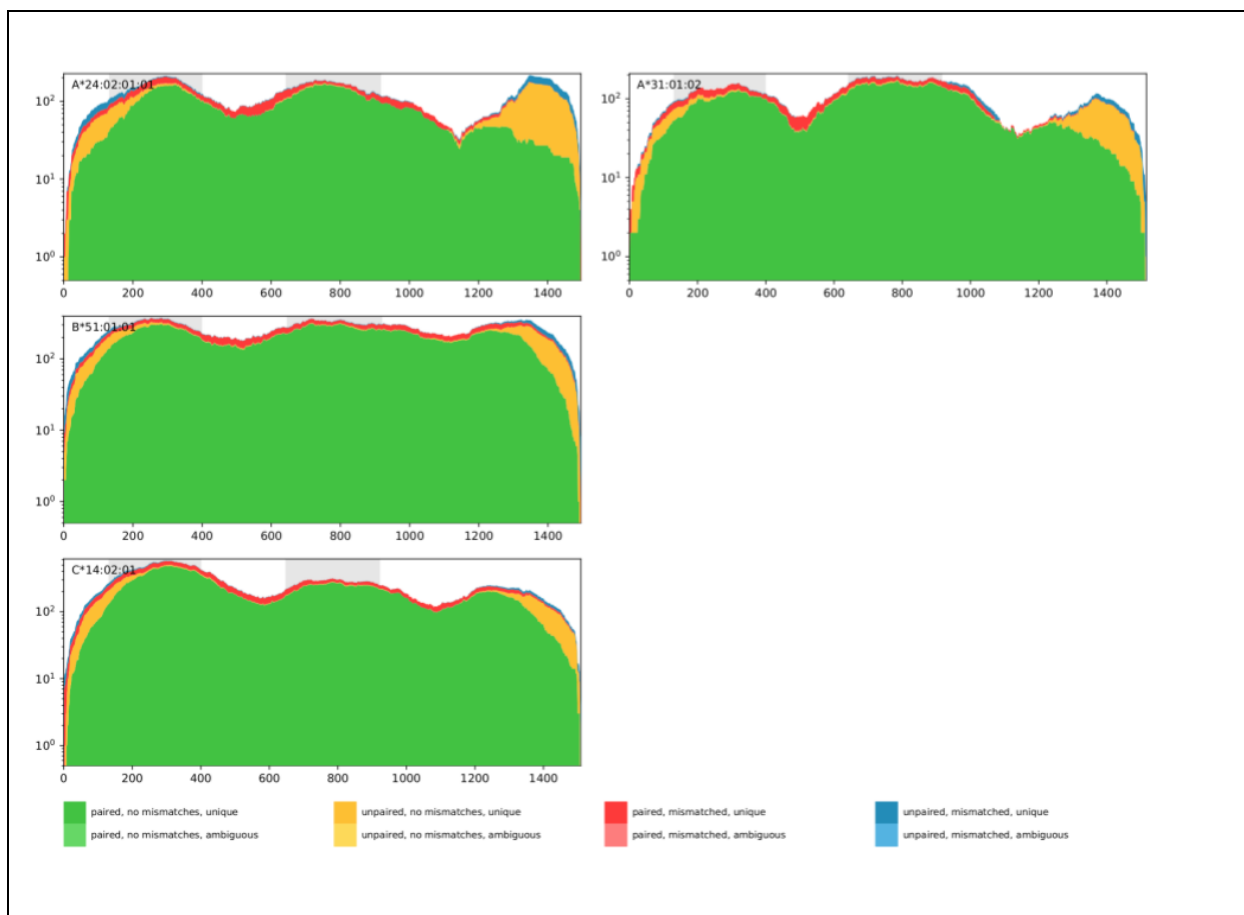


Figura S2-1: Resultado del alineamiento para los alelos identificados como los más probables para el HLA clase I de la paciente. En el eje X se representa la longitud del alelo. En el eje Y se representa la cantidad de lecturas mapeadas.

Anexo 3: Resultados detallados de la predicción de epítopes

Secuencia larga MT	Secuencia corta MT	Secuencia corta WT	Longitud péptido corto	Longitud péptido largo	Alelos que intersecta	FPK M	IC50[nm] MT respecto a los alelos anteriores	IC50[nm] WT respecto a los alelos anteriores	Vida media en horas respecto a los alelos anteriores
RPLWAW FQ RARVQLRRDLR	FQ RARVQLR	FQ RARVSSS	9	19	HLA-A*31:01 HLA-DPA1*02:02-DPB1*100:01 HLA-DPA1*02:02-DPB1*05:01	180	51.06 / 7 / 186	16630.9 / 50 / 1666	3.32 => sólo hay valores para clase I
KPYIY AGY R PLY Q TW RFYFR T	GYRPLYQ TW	GYRPLHQ TW	9	21	HLA-A*24:02 HLA-A*31:01 HLA-DPA1*02:02-DPB1*100:01 HLA-DPA1*02:02-DPB1*05:01	20	174 / 215 / 166 / 174	368 / 335 / 414 / 403	1.28 / 1.1 => sólo hay valores para clase I
YAGYR PLY Q TW R FY FR TL	LYQTWR FY F	LHQ TW R FY F	9	18	HLA-A*24:02 HLA-A*31:01 HLA-C*14:02 HLA-DPA1*02:02-DPB1*100:01 HLA-DPA1*02:02-DPB1*05:01	20	15.57 / 464.16 / 58.81 / 138.72 / 174.12	348.46 / 3758.78 / 2664.09 / 341.46 / 403.72	31.09 / 0.22 / 0.13 => sólo hay valores para clase I
LAGTL GR MLAV TFL LQA	RMLAV TFL L	SMLAV TFL L	9	17	HLA-A*24:02 HLA-A*31:01 HLA-DPA1*02:02-DPB1*100:01 HLA-DPA1*02:02-DPB1*05:01	50	97.2 / 130.59 / 20.22 / 401.78	331.98 / 595.55 / 100 / 950.07	3.63 / 0.47 => sólo hay valores para clase I
NA	ISKKKHYNR	ISKKKALQQ	9	NA	HLA-A*31:01	8	8.6	21837	5.48 => sólo hay valores para clase I
NA	RQQQ QHL HR	Q Q QQ QHL HR	9	NA	HLA-A*31:01	34	34.4	449.3	6.65 => sólo hay valores para clase I
NA	RQQQ QIL R	Q Q QQ QIL R	9	NA	HLA-A*31:01	44	44.1	614.9	4.6 => sólo hay valores para clase I
NA	RMRMS STPP R	NA	10	NA	HLA-A*31:01	8	8	NA	10.62 => sólo hay valores para clase I
YAGYR PLY Q TW R FY FR TL FQ	PLYQTWR FY F	PLHQ TW R FY F	10	20	HLA-A*24:02 HLA-A*31:01 HLA-DPA1*02:02-DPB1*05:01 HLA-DPA1*02:02-DPB1*100:01	20	280 / 74 / 178 / 104	1712 / 124 / 345 / 248	0.4 / 1.52 => sólo hay valores para clase I
CKDL IKF MLR NER Q F KEEKL	KFMLR NER Q F	K S MLR NER Q F	10	20	HLA-A*24:02 HLA-A*31:01 HLA-DPA1*02:02-DPB1*05:01 HLA-DPA1*02:02-DPB1*100:01	25	407 / 196 / 318 / 350	4701 / 256 / 874 / 1146	1.32 / 1.6 => sólo hay valores para clase I
NA	LWARRCP WT M	NA	10	NA	HLA-A*31:01	476	126	NA	0.52 => sólo hay valores para clase I
NA	HFKAPK ISMR	HFKAPK ISMP	10	NA	HLA-A*31:01	37	20	23453	0.75 => sólo hay valores para clase I

Tabla S3-1: Lista de péptidos candidatos a vacuna (MT: Mutant, WT: WildType).

Anexo 4: Resultados del *docking* molecular



Figura S4-1: Vista superior del bolsillo HLA. En naranja se presenta el péptido GYRPLHQTW. Las líneas azul celeste corresponden a los puentes de hidrógeno.

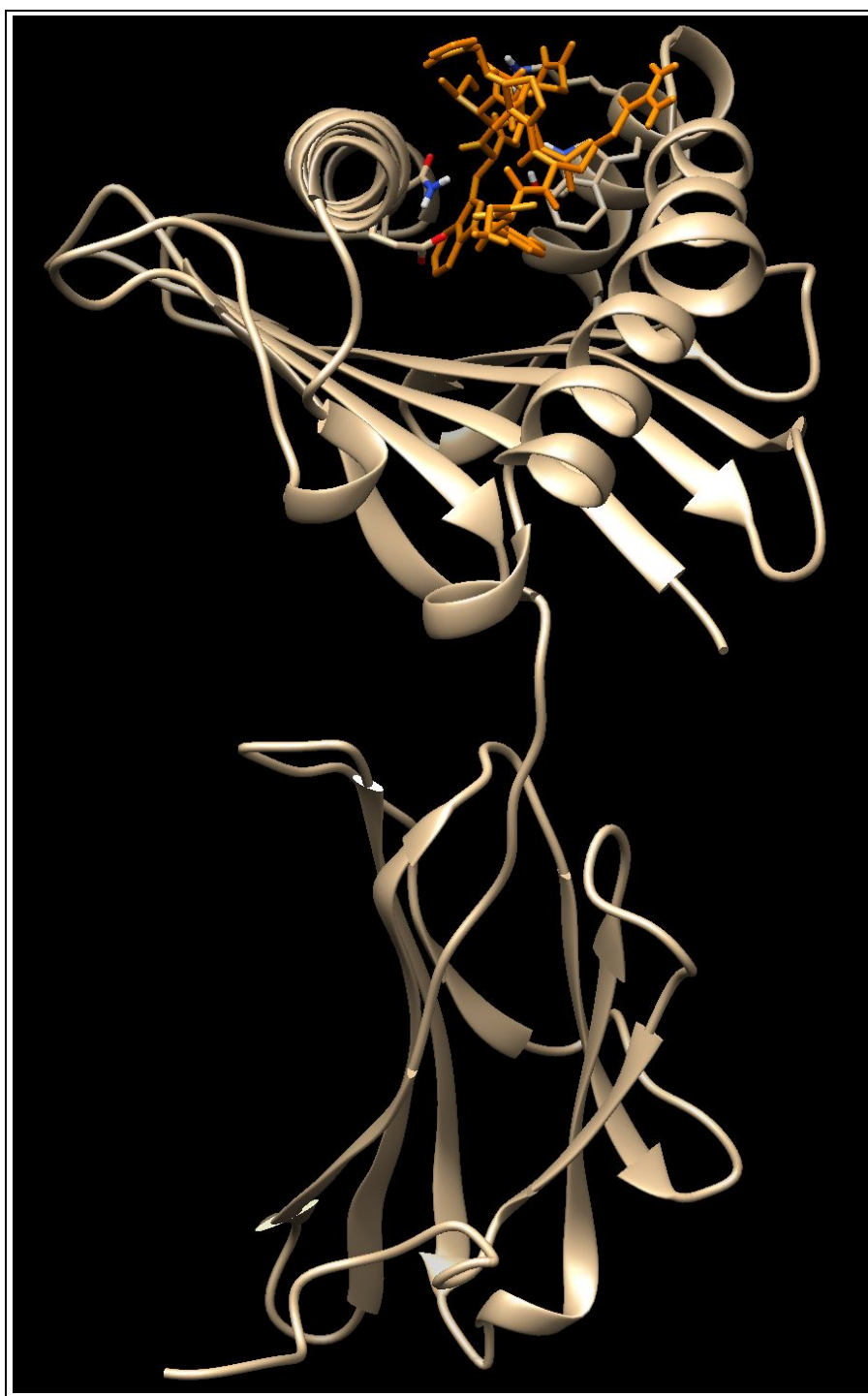


Figura S4-2: Vista lateral del bolsillo HLA. En naranja se presenta el péptido GYRPLHQTW.

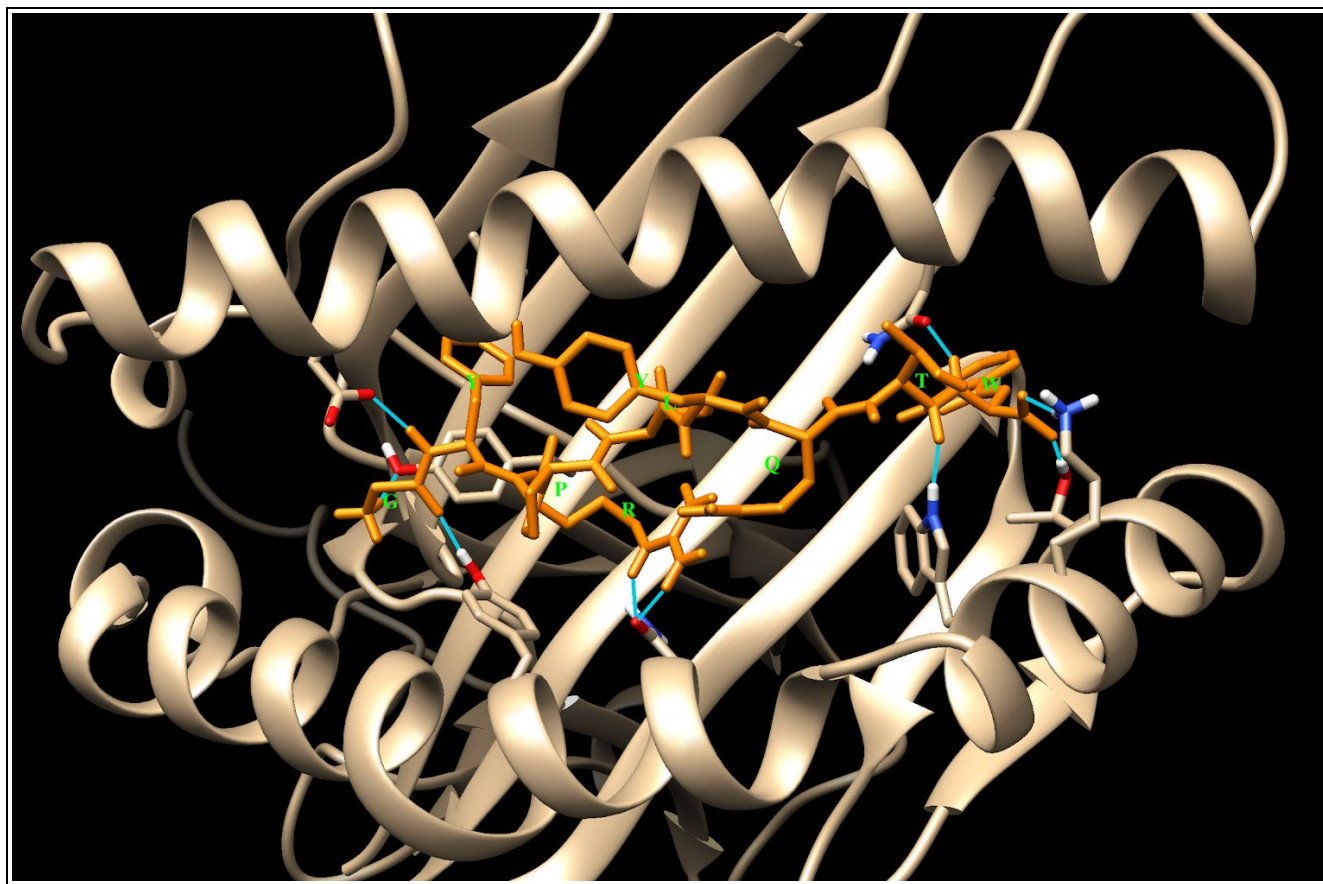


Figura S4-3: Vista superior del bolsillo HLA. En naranja se presenta el péptido GYRPLYQTW. Las líneas azul celeste corresponden a los puentes de hidrógeno.



Figura S4-4: Vista lateral del bolsillo HLA. En naranja se presenta el péptido GYRPLHQTW.

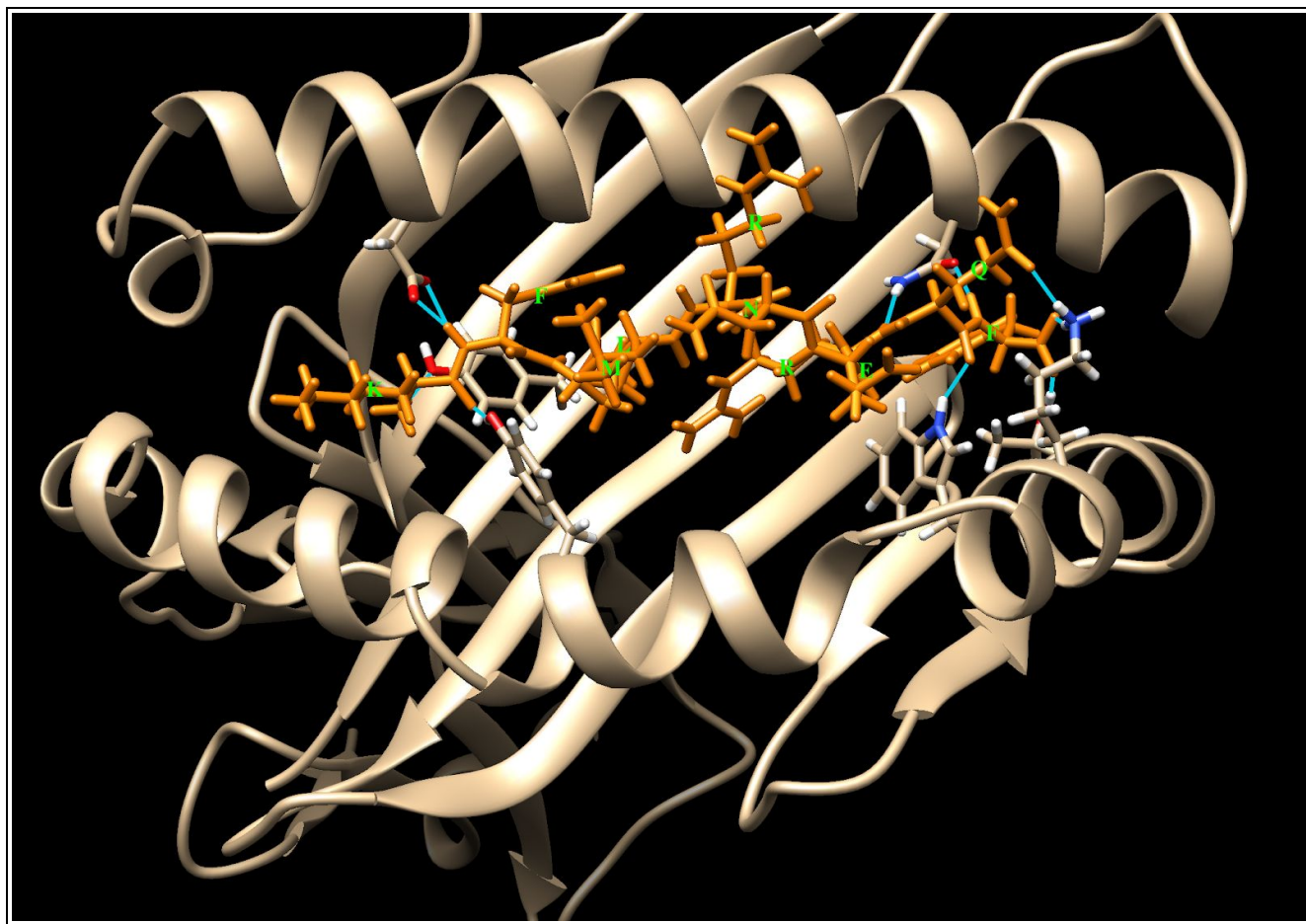


Figura S4-5: Vista superior del bolsillo HLA. En naranja se presenta el péptido KFMLRNERQF. Las líneas azul celeste corresponden a los puentes de hidrógeno.



Figura S4-6: Vista lateral del bolsillo HLA. En naranja se presenta el péptido KFMLRNERQF.

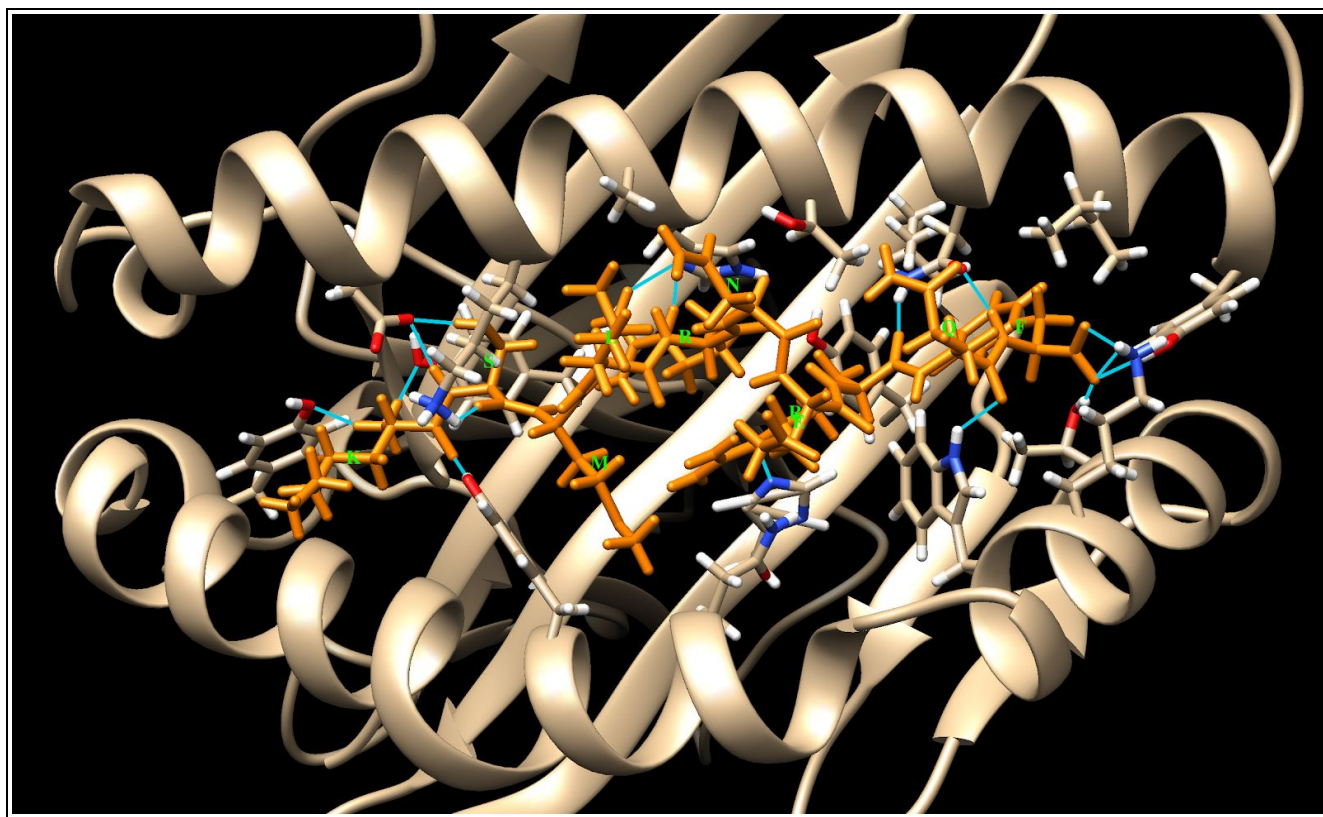


Figura S4-7: Vista superior del bolsillo HLA. En naranja se presenta el péptido KSMLRNERQF. Las líneas azul celeste corresponden a los puentes de hidrógeno.



Figura S4-8: Vista lateral del bolsillo HLA. En naranja se presenta el péptido KSMLRNERQF.

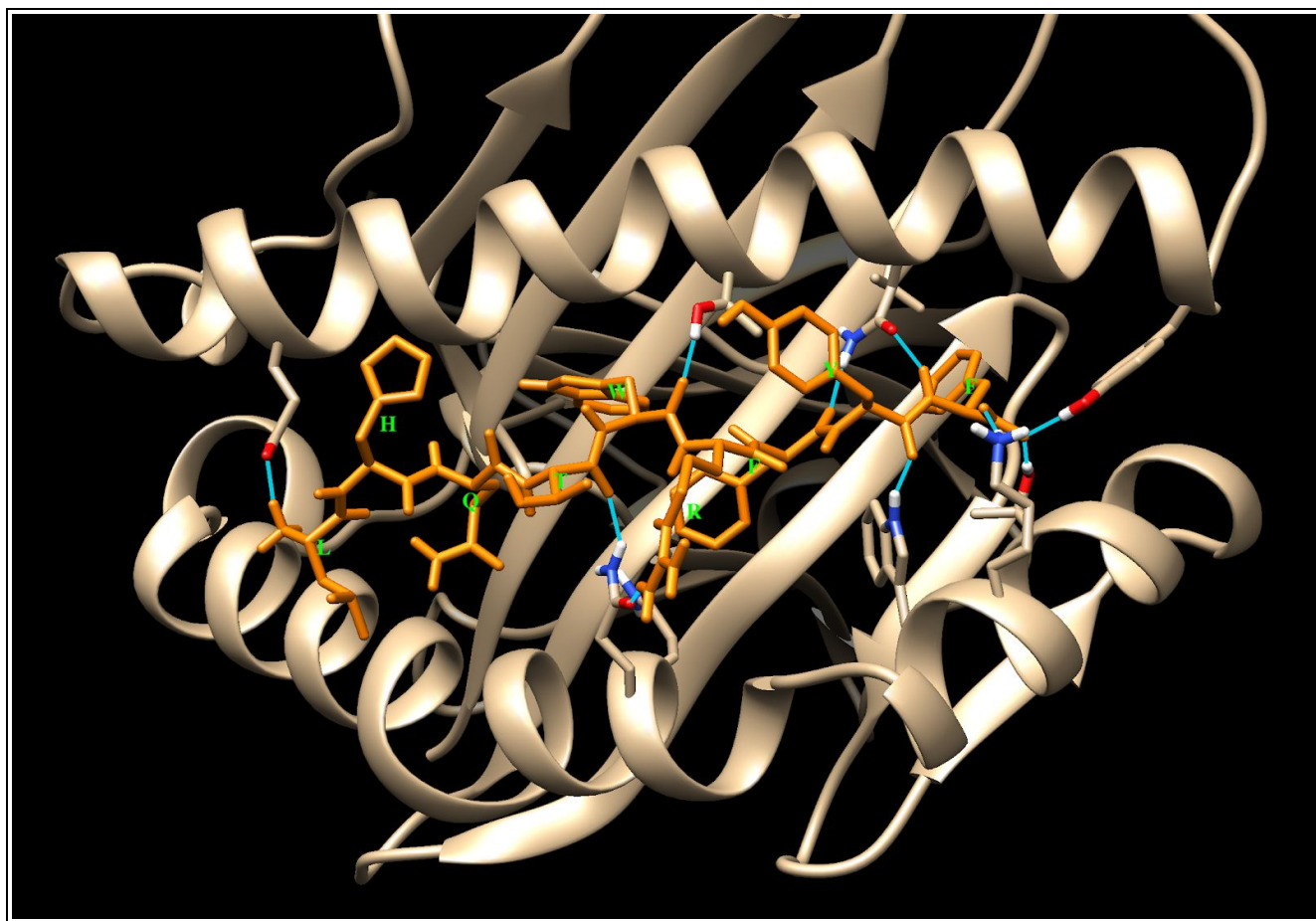


Figura S4-9: Vista superior del bolsillo HLA. En naranja se presenta el péptido LHQTWRFYF. Las líneas azul celeste corresponden a los puentes de hidrógeno.

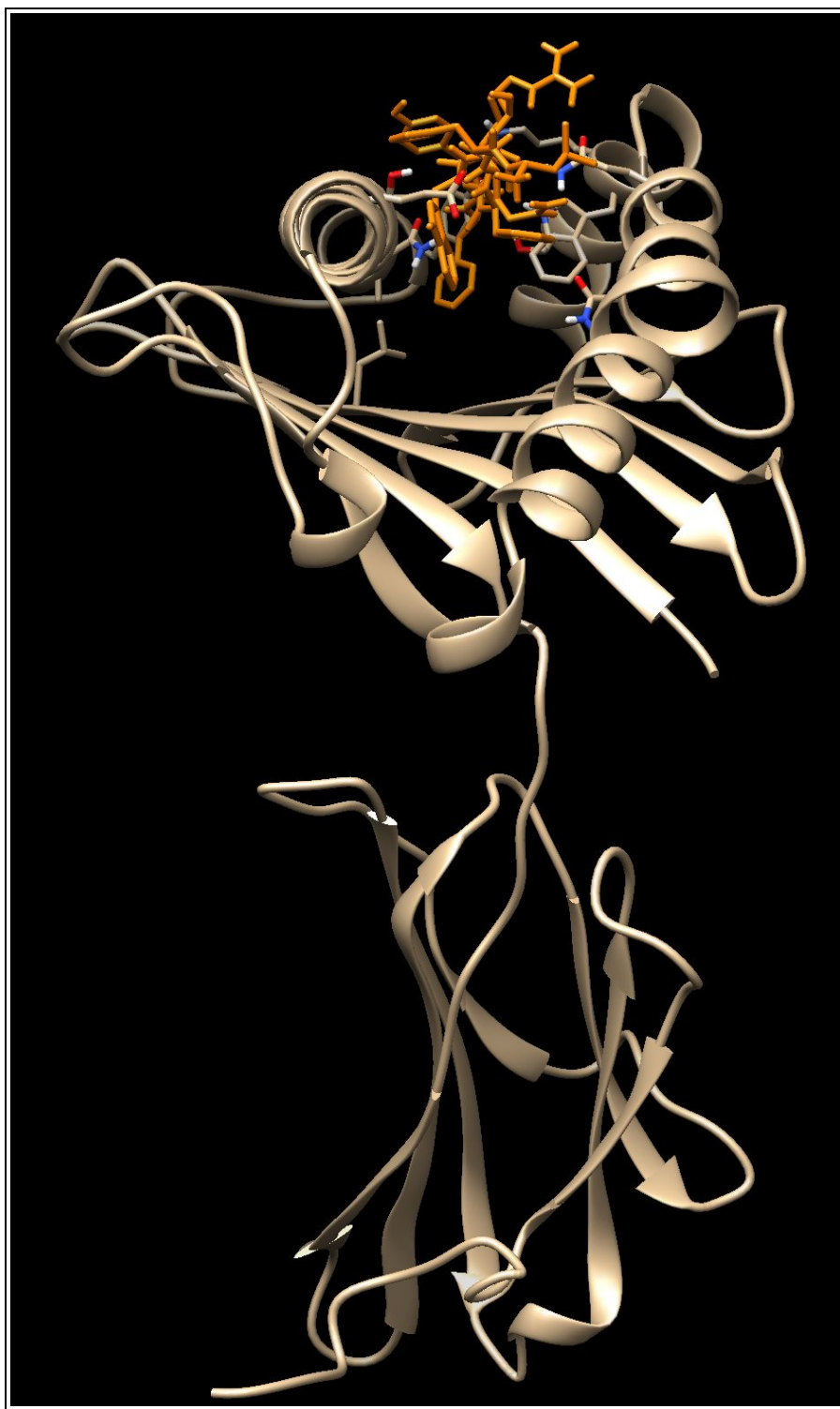


Figura S4-10: Vista lateral del bolsillo HLA. En naranja se presenta el péptido LHQTWRFYF.

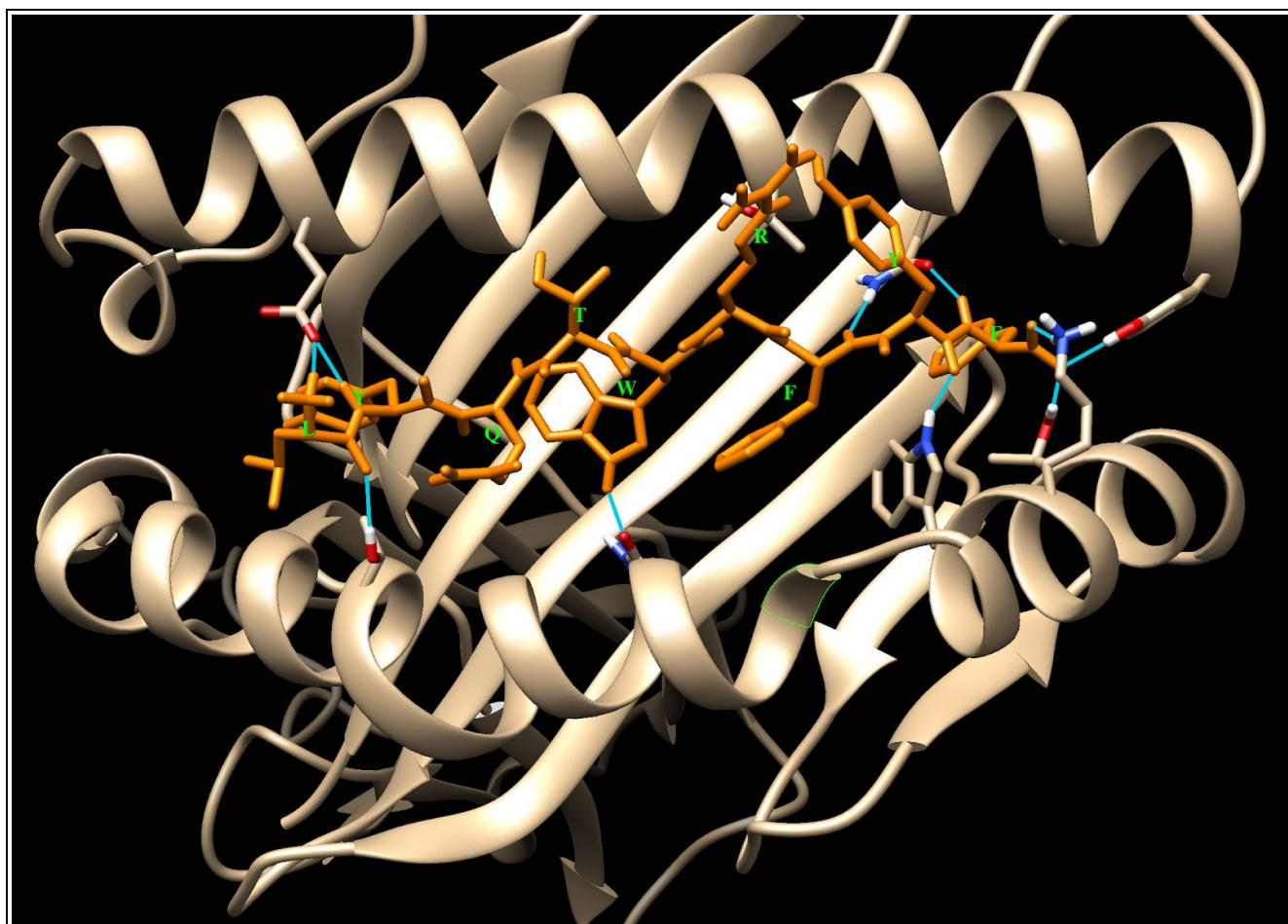


Figura S4-11: Vista superior del bolsillo HLA. En naranja se presenta el péptido LYQTWRFYF. Las líneas azul celeste corresponden a los puentes de hidrógeno.

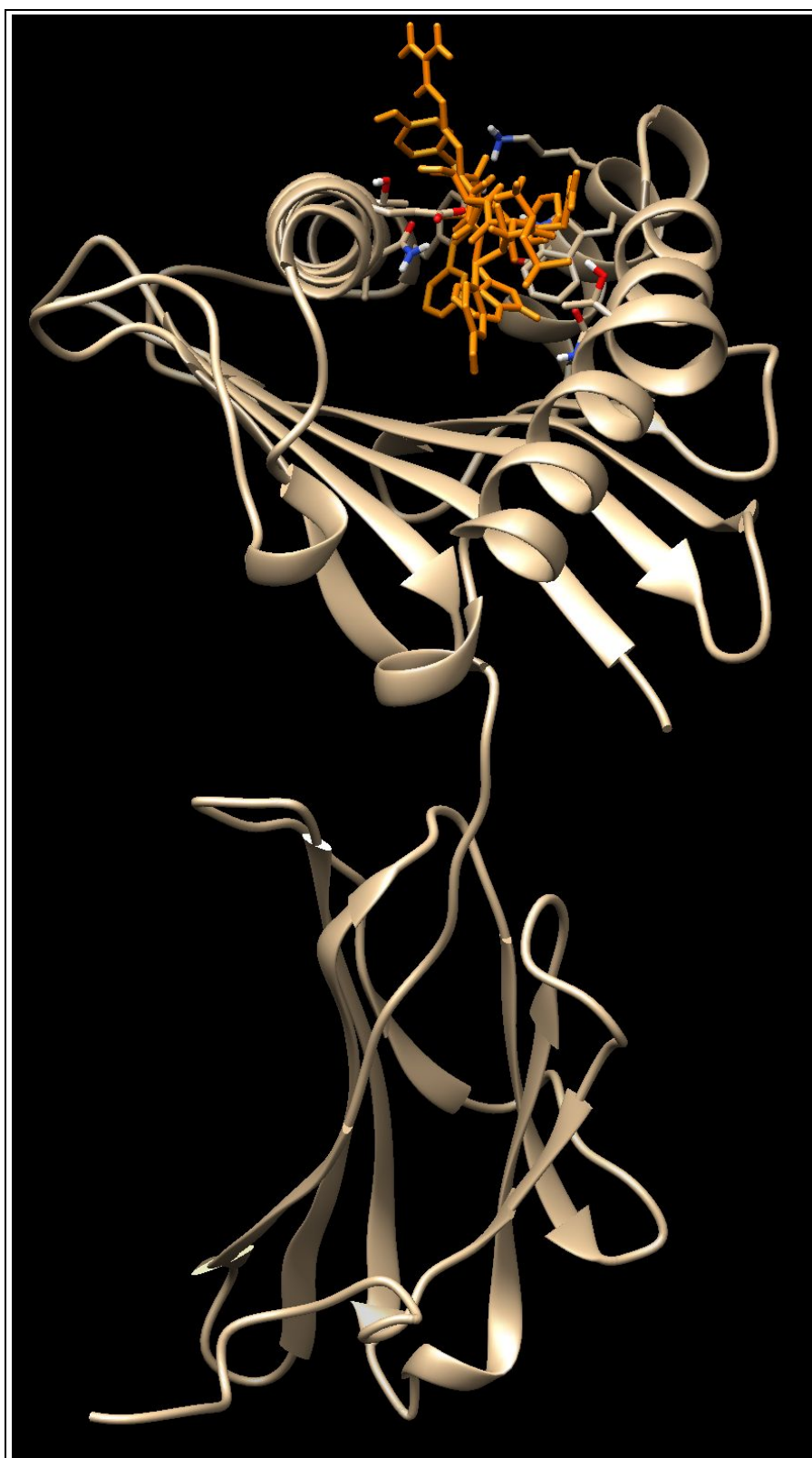


Figura S4-12: Vista lateral del bolsillo HLA. En naranja se presenta el péptido LYQTWRFYF.

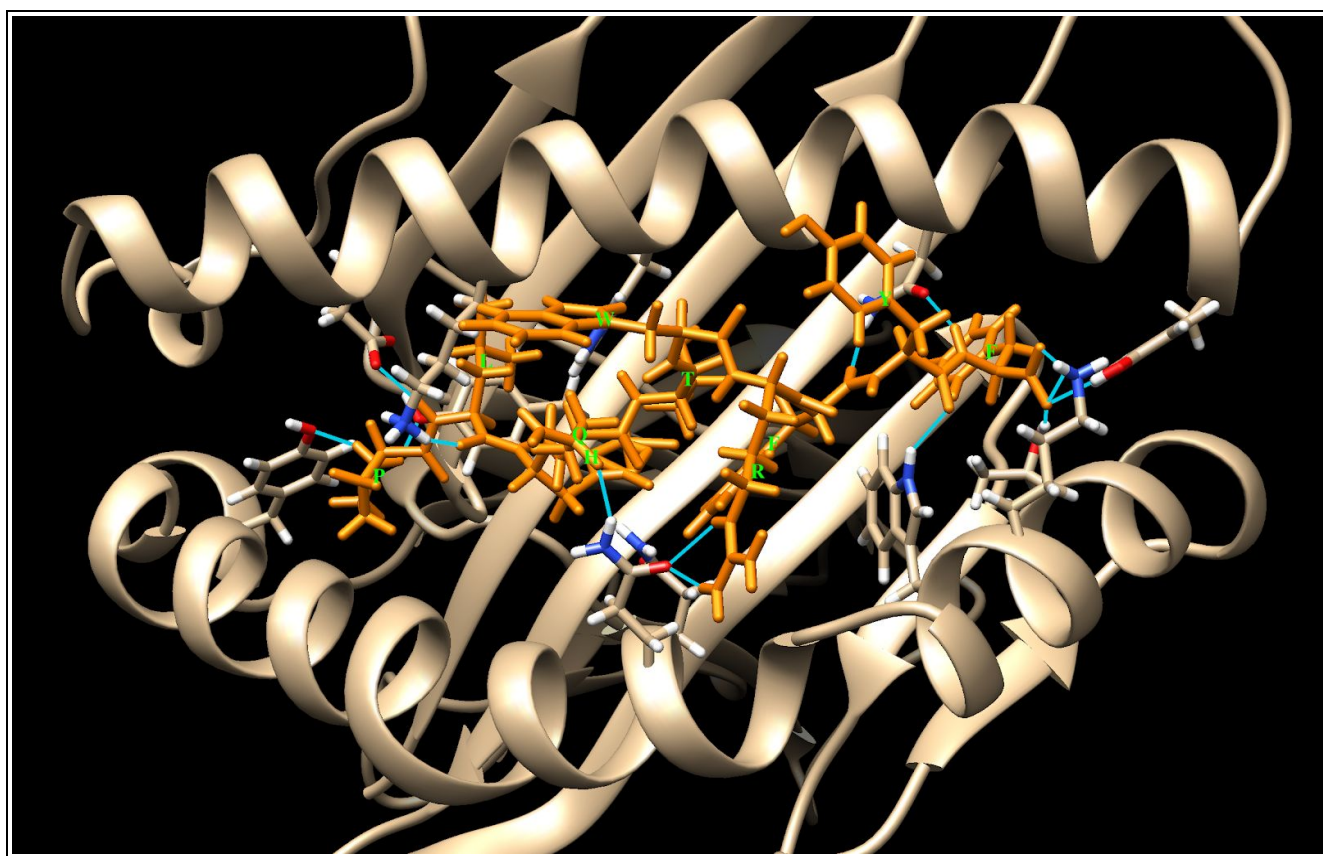


Figura S4-13: Vista superior del bolsillo HLA. En naranja se presenta el péptido PLHQTWRFYF. Las líneas azul celeste corresponden a los puentes de hidrógeno.



Figura S4-14: Vista lateral del bolsillo HLA. En naranja se presenta el péptido PLHQTWRFYF.

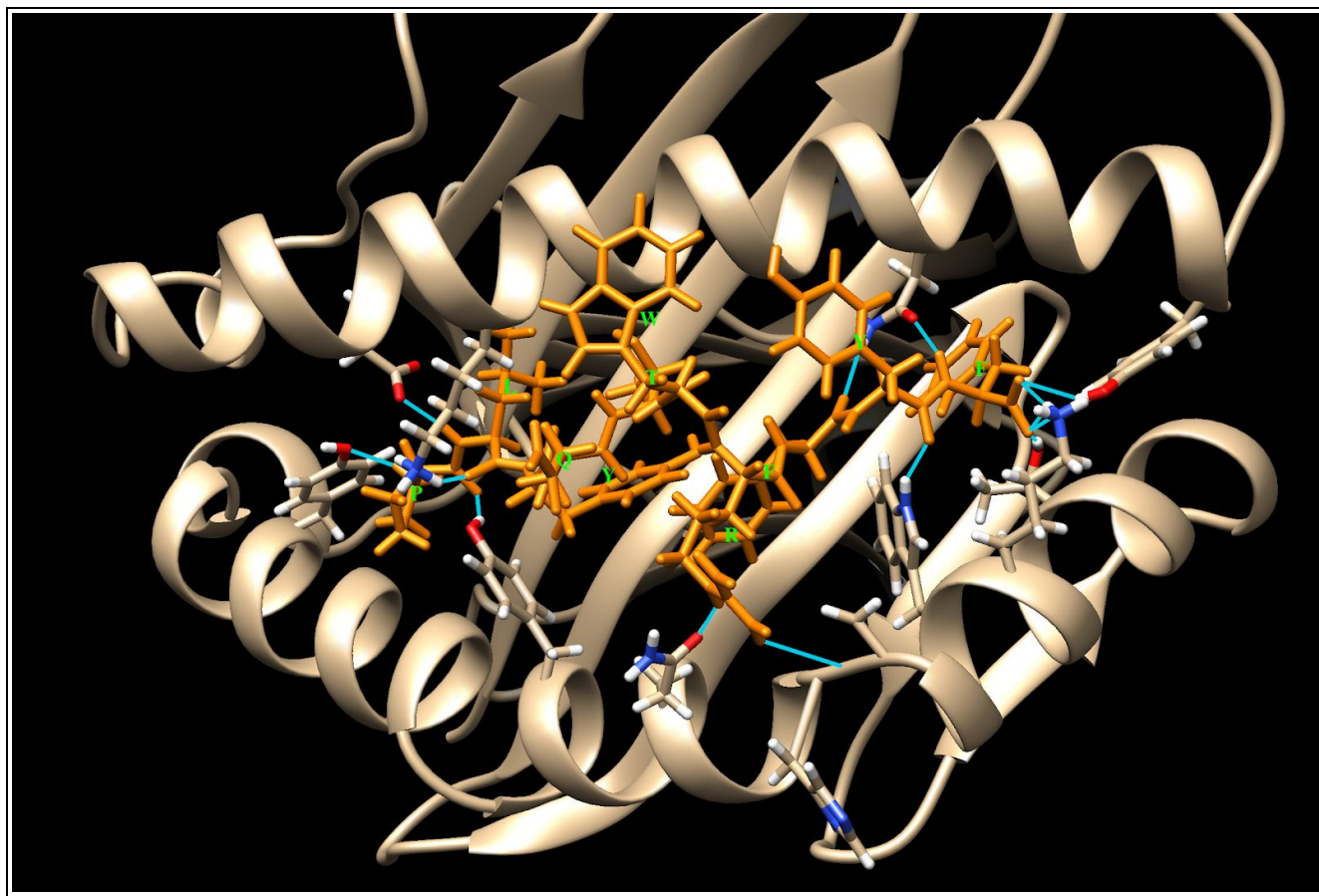


Figura S4-15: Vista superior del bolsillo HLA. En naranja se presenta el péptido PLYQTWRFYF. Las líneas azul celeste corresponden a los puentes de hidrógeno.



Figura S4-16: Vista lateral del bolsillo HLA. En naranja se presenta el péptido PLYQTWRFYF.

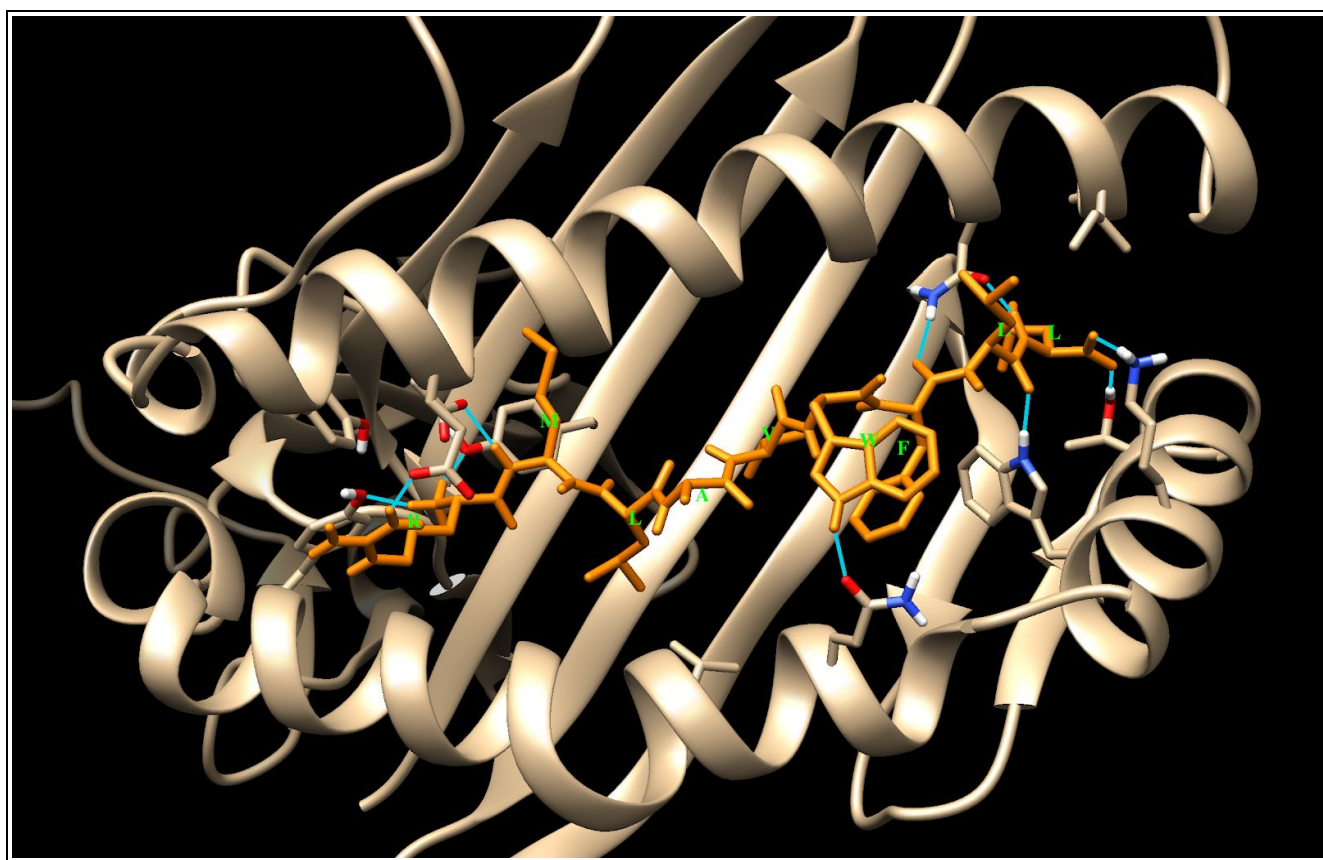


Figura S4-17: Vista superior del bolsillo HLA. En naranja se presenta el péptido RMLAVTFLL. Las líneas azul celeste corresponden a los puentes de hidrógeno.



Figura S4-18: Vista lateral del bolsillo HLA. En naranja se presenta el péptido RMLAVTFLL.

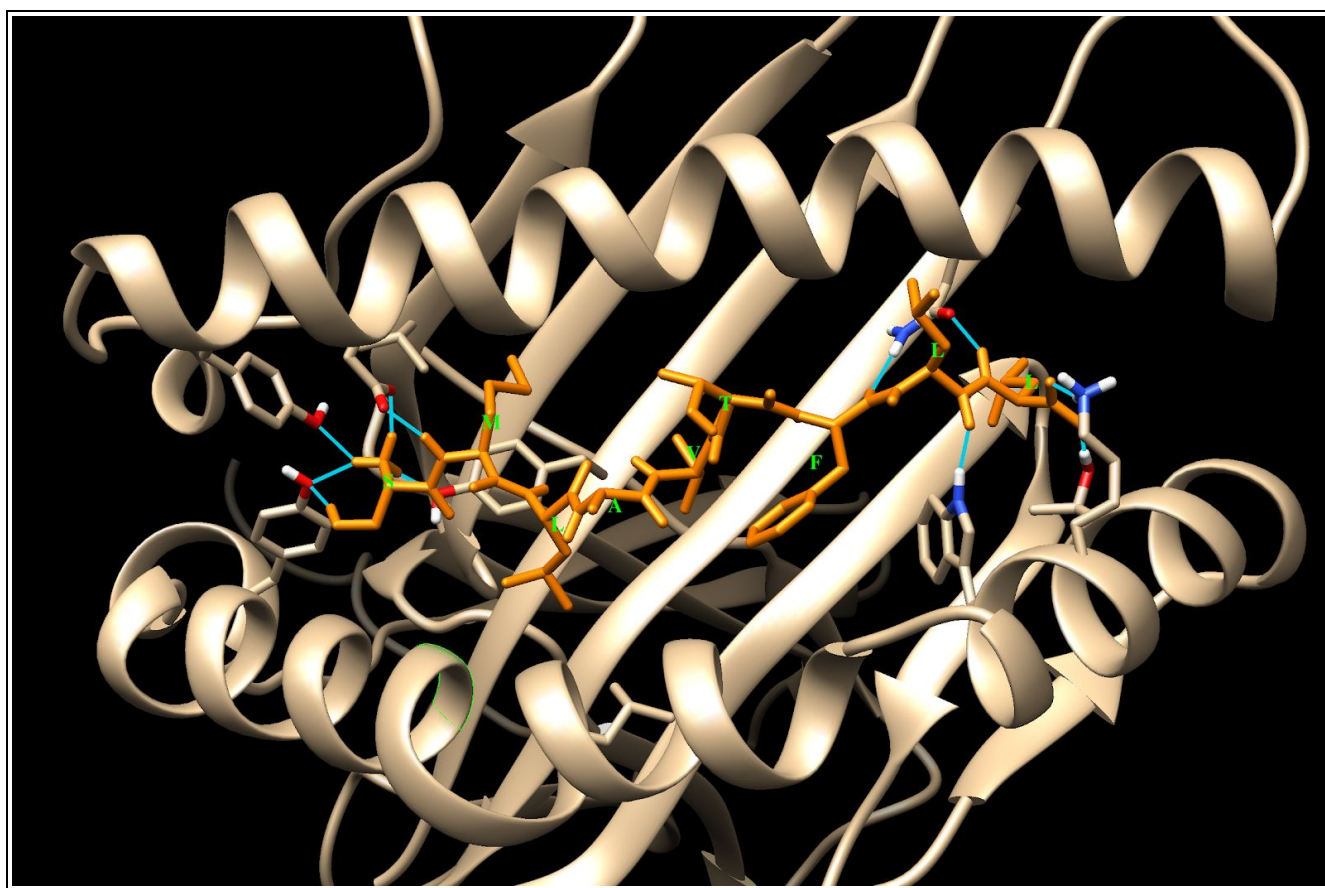


Figura S4-19: Vista superior del bolsillo HLA. En naranja se presenta el péptido SMLAVTFLL. Las líneas azul celeste corresponden a los puentes de hidrógeno.



Figura S4-20: Vista lateral del bolsillo HLA. En naranja se presenta el péptido SMLAVTFLL.